ILMATIETEEN LAITOS
METEOROLOGISKA INSTITUTET
FINNISH METEOROLOGICAL INSTITUTE

153
CONTRIBUTIONS

OBSERVING AND FORECASTING ROAD SURFACE TEMPERATURES

VIRVE KARSISTO

FINNISH METEOROLOGICAL INSTITUTE
CONTRIBUTIONS

No. 153

OBSERVING AND FORECASTING ROAD SURFACE TEMPERATURES

Virve Karsisto

Department of Physics
Faculty of Science
University of Helsinki
Helsinki, Finland

ACADEMIC DISSERTATION in meteorology

To be presented, with the permission of the Faculty of Science of the University of Helsinki, for public criticism in E204 auditorium at Physicum (Gustaf Hällströmin katu 2 A, Helsinki) on November 1st, 2019, at 12 noon.

Finnish Meteorological Institute
Helsinki, 2019

Supervisors Professor Heikki Järvinen
Department of Physics
University of Helsinki, Finland

Lic.Phil. Pertti Nurmi
Meteorological Research, Meteorological research applications
Finnish Meteorological Institute, Finland

Dr. Timo Sukuvaara
Space Research and Observation Technologies, Observation and Measuring Systems
Finnish Meteorological Institute, Finland

Reviewers Professor Thorsten Cypra
School of Architecture and Civil Engineering
University of Applied Sciences, Saarbrücken, Germany

Professor Andreas Christen
Faculty of Environment and Natural Resources
University of Freiburg, Germany

Custos Professor Heikki Järvinen
Department of Physics
University of Helsinki, Finland

Opponent Professor Lee Chapman
School of Geography, Earth and Environmental Sciences
University of Birmingham, UK

ISBN 978-952-336-078-5 (paperback)
ISBN 978-952-336-079-2 (pdf)
ISSN 0782-6117

Edita Prima Oy
Helsinki 2019

ILMATIETEEN LAITOS
METEOROLOGISKA INSTITUTET
FINNISH METEOROLOGICAL INSTITUTE

Published by Finnish Meteorological Institute
(Erik Palménin aukio 1), PL 503
00101 Helsinki

Series title, number and report code of publication
Finnish Meteorological Institute
Contributions 153, FMI-CONT-153
Date
November 2019

Author
Virve Karsisto
ORCID iD https://orcid.org/0000-0002-5212-1002

Title

Observing and forecasting road surface temperatures

Abstract

Wintertime weather conditions can be hazardous for road traffic. Icy roads and poor visibility caused by snowfall increase the accident risk. Accurate forecasting of road conditions is important, because reliable and precise forecasts help the road maintenance personnel to plan their operations accordingly. Well timed maintenance operations increase safety and enable economical savings as unnecessary actions can be avoided. Drivers can also adjust their route plan and driving behaviour appropriately when warnings of hazardous conditions are given well beforehand.

Road conditions are forecasted in the Finnish Meteorological Institute (FMI) with specialized road weather model. Before executing the actual forecast, the model is first initialized by feeding it with observation data. The quality of this data is essential for forecast accuracy, as the forecast is greatly dependent of the initial model state. Road weather stations have traditionally been one of the main sources of information, but their density is sparse especially in rural areas. Road surface temperature can vary considerably across the road network, so observations should be done in dense enough spatial scale. Nowadays it is possible to gather real time information from vehicles. Mobile sources provide observations with high spatial density and thus facilitate detecting the road stretches most prone to freezing. However, the quality of mobile observations should be assessed before implementing them to the road weather forecasting systems.

This dissertation aims to answer to two research questions. Firstly, it has been studied how to best use available surface temperature observations in the road weather model initialization. Secondly , it has been studied how differences in two road weather models' physics affect to the surface temperature forecast accuracy.

A method called coupling was implemented to the FMI road weather model. The main idea of the method is to adjust the incoming radiation flux so that the modelled surface temperature fits to the last observed value. The results show that this method improves considerably the short range surface temperature forecasts.

Mobile surface temperature observations done with Teconer RTS411 were compared to road weather station measurements to assess the mobile data quality. According to the results, the mobile observations were on average 0.62 °C warmer than the road weather station measurements at 0 °C and in dry conditions. It was found out that the difference between mobile observations and road weather station measurements was dependent on the road status. A calibration equation for mobile observations was developed using linear mixed models to get mobile observations more in line with road weather station measurements.

The effect of the mobile observations to the road surface temperature forecast accuracy was studied. According to the results, using the mobile observations calibrated with the developed equation improved the accuracy of road surface temperature forecasts compared to a theoretical situation where there would not be other surface temperature observations available. However, for an area with a dense road weather station network the accuracy of forecasts assimilating mobile observations with correction were on par with the accuracy of forecast assimilating interpolated surface temperature values.

Studying model physics and comparing behaviour of different models is beneficial for model development. In this work, the verification results of the FMI's and the Royal Netherlands Meteorological institute's (KNMI) road weather models were compared to each other. In addition, the model physics were studied to find out the reasons for differences in the surface temperature forecasts. The forecasts of the KNMI model were found to be slightly more accurate than the forecasts of the FMI model. Although the core physics of the models were rather similar, there were large differences in some physical parameters and the number and the thickness of the ground layers. Individual reason for the better performance of the KNMI model could not be found, as the effects of different physical properties eventually sum up to surprisingly similar modelled surface temperature values.

Publishing unit
Finnish Meteorological Institute, Meteorological research

Classification (UDC)
551.501, 551.509
551.525, 551.52

Keywords
Road surface temperature, mobile observations, energy balance model, road weather forecast, model comparison, forecast verification

ISSN and series title
0782-6117 Finnish Meteorological Institute Contributions

ISBN
978-952-336-078-5 (paperback)
978-952-336-079-2 (pdf)

DOI
https://doi.org/10.35614/isbn.9789523360792

Language
English

Pages
116

ILMATIETEEN LAITOS
METEOROLOGISKA INSTITUTET
FINNISH METEOROLOGICAL INSTITUTE

Julkaisija
Ilmatieteen laitos
(Erik Palménin aukio 1), PL 503
00101 Helsinki

Julkaisun sarja, numero ja raporttikoodi
Finnish Meteorological Institute
Contributions 153, FMI-CONT-153
Julkaisuaika
Marraskuu 2019

Tekijä
Virve Karsisto
ORCID iD https://orcid.org/0000-0002-5212-1002

Nimike
Tienpintalämpötilan havainnointi ja ennustaminen

Tiivistelmä

Talviset sääolosuhteet, kuten jäiset tiet ja lumisateen aiheuttama heikko näkyvyys, lisäävät onnettomuuksien riskiä. Teiden turvallisena pitämisessä auttavat tarkat ja luotettavat tiesääennusteet, joiden avulla auraukset ja suolaukset voidaan suunnitella hyvissä ajoin. Tarkat ennusteet mahdollistavat myös taloudelliset säästöt, kun tarpeettomia toimenpiteitä pystytään välttämään. Lisäksi tienkäyttäjät voivat suunnitella reittinsä olosuhteet huomioon ottaen, kun varoitukset huonosta kelistä annetaan hyvissä ajoin.

Ilmatieteen laitos ennustaa keliolosuhteita kyseistä tarkoitusta varten kehitetyn tiesäämallin avulla. Ennen varsinaisen ennusteen tekemistä malli alustetaan havainnoilla. Ennuste on hyvin riippuvainen mallin alkutilasta, joten havaintodatan laatu on ennusteen tarkkuuden kannalta tärkeää. Perinteisesti havaintojen tärkein lähde ovat olleet tiesääasemat, mutta niitä on harvassa erityisesti Pohjois-Suomessa. Tienpintalämpötila voi vaihdella merkittävästi tiestön eri osissa, joten havaintoja pitäisi tehdä riittävän tiheästi. Nykyisin on mahdollista saada reaaliaikaisia havaintoja myös autoista. Autohavaintoja voidaan tehdä paljon tiheämmin kuin tiesääasemahavaintoja, jolloin myös kaikkein herkimmin jäätyvät tiestön osat havaitaan helpommin. Autohavaintojen laatu pitää kuitenkin varmistaa ennen kuin niitä voidaan käyttää tiesäämallinnuksessa.

Tämä väitöskirjatyö pyrkii vastaamaan kahteen tutkimuskysymykseen. Ensinnäkin tutkittiin, miten pintalämpötilahavaintoja voidaan parhaiten käyttää tiesäämallin alkutilan määrityksessä. Lisäksi tutkittiin, miten erot kahden eri tiesäämallin fysiikassa vaikuttavat tienpintalämpötilaennusteen tarkkuuteen.

Ilmatieteen laitoksen tiesäämalliin implementoitiin ”coupling”nimellä kulkeva menetelmä, jonka avulla säteilyvuota säädetään mallissa siten, että mallin pintalämpötila vastaa viimeisintä havaintoa. Saatujen tulosten perusteella menetelmä paransi huomattavasti lyhyen aikaskaalan tienpintalämpötilaennusteita.

Mobiilimittausten laatua selvitettiin vertaamalla autoon kiinnitettävällä Teconer RTS411 laitteella tehtyjä tienpintalämpötilamittauksia tiesääasemien mittauksiin. Tulosten mukaan mobiilihavainnot olivat keskimäärin 0.62 °C tiesääasemahavaintoja lämpimämpiä, kun tie oli kuiva ja lämpötila nollassa. Tutkimuksessa havaittiin, että ero mobiilien ja tiesääasemamittausten välillä oli riippuvainen siitä, oliko tienpinta kuiva vai esimerkiksi märkä tai jäinen. Tutkimuksessa kehitettiin kalibrointiyhtälö mobiilihavaintojen korjaamiseksi lineaarisia sekamalleja hyödyksi käyttäen.

Mobiilihavaintojen vaikutusta tienpintalämpötilaennusteen tarkkuuteen tutkittiin vertaamalla mobiileja tienpintalämpötilahavaintoja hyödyntävää ennustetta kahteen kontrolliennusteeseen. Kalibroitujen mobiilihavaintojen käyttö paransi ennusteita verrattuna teoreettiseen tilanteeseen, jossa tienpintalämpötilahavaintoja ei ollut saatavilla. Sen sijaan mobiilihavaintoja käyttävät ennusteet antoivat suunnilleen yhtä tarkkoja ennusteita kuin interpoloituja tienpintalämpötilahavaintoja käyttävä malli ennustealueella, jolla on tiesääasemia tiheässä.

Mallifysiikan tutkiminen ja eri mallien käyttäytymisen vertailu on hyödyllistä mallikehitykselle. Tässä työssä verrattiin Suomen Ilmatieteen laitoksen ja Alankomaiden Ilmatieteen laitoksen tiesäämallien ennustetuloksia keskenään. Myös mallien fysikaalisia ominaisuuksia verrattiin, jotta saataisiin selville, mistä erot tienpintalämpötilaennusteissa johtuvat. Alankomaiden mallin tienpintalämpötilaennusteiden havaittiin olevan hieman tarkempia kuin Suomen Ilmatieteen laitoksen mallin. Vaikka perusfysiikka oli melko samankaltainen molemmissa malleissa, joittenkin fysikaalisten parametrien arvoissa oli suuria eroja. Mallien maakerrosten lukumäärä ja paksuus erosivat myös toisistaan. Yksittäistä syytä sille, miksi Alankomaiden malli tuotti hieman tarkempia ennusteita, ei löydetty. Fysikaalisista eroista huolimatta eri tekijät malleissa summautuvat siten, että mallien ennustama pintalämpötila oli huomattavan samanlainen.

Julkaisijayksikkö
Ilmatieteen laitos, Meteorologinen tutkimus

Luokitus (UDK)	Asiasanat	
551.501, 551.509 551.525, 551.52	Tienpintalämpötila, mobiilihavainnot, energiatasapainomalli, tiesääennuste, mallivertailu, ennusteiden verifiointi	
ISSN ja avainnimike	ISBN	
0782-6117 Finnish Meteorological Institute Contributions	978-952-336-078-5 (nid.), 978-952-336-079-2 (pdf)	
DOI	Kieli	Sivumäärä
https://doi.org/10.35614/isbn.9789523360792	Englanti	116

Preface

I have done my dissertation work in Meteorological Research Applications group in the Finnish Meteorological Institute. I want to thank my first group leader and supervisor Pertti Nurmi for taking me in as a master student and for his support and positive feedback. Special thanks to my two other supervisors Heikki Järvinen and Timo Sukuvaara for their support. Many thanks also to the unit leader Sylvain Joffre for his help and comments during the time when I was working with my first article. Regarding the other articles in my thesis I am grateful especially for Lauri Lovén for his hard work with the third paper. I want to thank also Sander Tijm for his co-authorship and support with my second article.

It might not have been easy to fill Pertti Nurmi's shoes, but our new group leader Anders Lindfors has done a great job so far. I want to thank him for his guidance and support in my work. I thank also all the colleagues who have helped me during my dissertation work. Especially I want to thank Carl Fortelius for the helpful discussions during the earlier phase of my dissertation work. I am grateful also to our unit leader Timo Vihma and division leader Sami Niemelä for their support and administrative work. I appreciate the positive atmosphere in our unit, as nice colleagues here make working much more pleasant. As life is not only about working, I am really happy about the Pokémon Go players in our institute. It has been great to meet people from different divisions and play together!

I thank Suomen Kulttuurirahasto (the Finnish Cultural Foundation) for funding for my dissertation project. I thank also all the participants and funders of the projects where I have been able to do research. Some of the main projects were Intelligent Arctic trucks and Sod5G funded by European Regional Development Fund of the European Union and Regional Council of Lapland and WiRMa funded by Interreg Nord fund of European Union and Regional Council of Lapland.

Lastly I want to thank my husband Petteri Karsisto, to whom I can share both the positive and negative things of my work.

Virve Karsisto
Helsinki, August 2019

Contents

Acronyms - Abbreviations 9
List of original publications 10
1. Introduction 11
2. Road weather model 15
2.1. Input and output 15
2.2 Surface temperature calculation 16
2.2.1 Heat transfer in the ground 16
2.2.2 Surface energy balance 20
2.2.3 Sensible heat flux 20
2.2.4 Latent heat flux 21
2.3 Relaxation 22
2.4 Coupling 22
2.4 Storage terms 24
3. Data 26
3.1 Observations 26
3.3.1 Road weather stations 26
3.3.2 Mobile measurements 26
3.2 Atmospheric forecasts 27
3.3 Kriging 27
4. Summary of the results 29
4.1 The effect of coupling and relaxation to short range forecasts 29
4.2 Comparison to the Netherlands RWM 30
4.3 The usability of mobile observations in road weather forecasts 33
5. Conclusions and future perspectives 37

References 39

Acronyms - Abbreviations

1D - One dimensional
3D - Three dimensional
AROME - the Application of Research to Operations at Mesoscale
FMI - Finnish Meteorological Institute
HARMONIE - Hirlam Aladin Research on Meso-scale Operational NWP in Europe
KNMI - Royal Netherlands Meteorological Institute
LT - Local time
LW - Long Wave
NWP - Numerical weather prediction model
METRo - Model of the Environment and Temperature of Roads
MSE - Mean squared error
RSME - Root mean squared error
RWM - Road weather model
RWS - Road weather station
SURFEX - Surface Externalisée (in French)
SW - Short Wave
UTC - Coordinated universal time

List of original publications

I Karsisto, V., P. Nurmi, M. Kangas, M. Hippi, C. Fortelius, S. Niemelä ja H. Järvinen, 2016: Improving road weather model forecasts by adjusting the radiation input. *Meteorological Applications*, **23**, 503–513, doi:10.1002/met.1574

II Karsisto, V., S. Tijm and P. Nurmi, 2017: Comparing the Performance of Two Road Weather Models in the Netherlands. *Weather and Forecasting*, **32**, no. 3, 991–1006, doi: 10.1175/WAF-D-16-0158.1

III Lovén L., V. Karsisto, H. Järvinen, M. J. Sillanpää, T. Leppänen, E. Peltonen, S. Pirttikangas, J. Riekki, 2019: Mobile road weather sensor calibration by sensor fusion and linear mixed models. *PLoS ONE* 14(2): e0211702, doi:10.1371/journal.pone.0211702

IV Karsisto V., Lovén L, 2019: Verification of road surface temperature forecasts assimilating data from mobile sensors. *Weather and Forecasting*, **34**, no. 3, 539–558, doi:10.1175/WAF-D-18-0167.1

The results of the four research articles listed above are summarized in this thesis. The author is responsible of the most of the data analysis and writing process related to papers **I**,**II** and **IV**. The author did the conceptual design related to papers **I** and **IV** and in paper **II** it was done together wit co-authors. In Paper **III**, the author performed the matching of mobile and stationary observations and participated in the writing process, data analysis and interpretation of the results.

1. Introduction

Slippery conditions cause problems to traffic every winter in countries such as Finland, where daily minimum temperatures often drop below zero. According to Nurmi et al. (2012), approximately 10% of wintertime accidents are induced by adverse weather in Finland. The estimated winter weather related accident costs are 226 million euros per winter. Their study included also a weather service chain analysis, which estimated that the available weather forecast information leads to the average savings of 37 million euros per winter. When the accuracy of weather forecasts increases, the savings would be even higher. In addition to the reduced accident costs, accurate forecasts lead to savings in the road maintenance. Salting and ploughing are expensive operations and the costs were approximately 1 000 euros per lane kilometer during year 2013 (Päiviö and Kärki, 2016). With accurate forecasts, maintenance operations can be well timed and unnecessary actions can be avoided. For example, melting ice when freezing has already occurred requires much more salt than preventing the freezing by salting the road beforehand (Thornes, 1991).

One of the most important aspects of road weather forecasts is to estimate when the road surface temperature drops below the freezing point. Equally important is to accurately forecast whether the road is wet at this point or is there enough humidity in the air to form black ice on the surface. Three-dimensional numerical weather prediction (NWP) models forecast air temperature, humidity, wind speed, radiation, precipitation and other weather variables over large areas. The grid size is usually from one kilometer to several kilometers. Additional road weather models (RWM) are used to estimate road conditions in detail. Many RWMs have been developed in the past years (Rayer, 1987; Jacobs and Raatz, 1996; Chapman and Thornes, 2011; Crevier and Delage, 2001; Fujimoto et al., 2012; Yang et al., 2012; Kangas et al., 2015). RWMs are typically one dimensional heat balance models that estimate temperature profile in the road. The RWM used in the Finnish Meteorological Institute (FMI) has been in operational use since 2000 (Kangas et al., 2015).

Road weather stations (RWS) measure road surface temperature and other variables along roads. They are essential source of information for determining the starting state in the RWM. However, installing RWSs is expensive and the station network is sparse in rural areas. Consequently, many road segments are left without measurements. There are many local features that affect the surface temperature, such as elevation and screening. Screening means that there is an object between the sun and the road location preventing the point from receiving direct solar radiation. Even 10 °C surface temperature differences can be detected in different parts of the road network (Shao et al., 1996; Bogren et al., 2000). These features should be taken into account in the RWMs to estimate accurately road surface temperature especially in the spots most vulnerable to freezing. During the past years several techniques are

presented to achieve this goal. They can be roughly divided into three categories: 1. Statistical/climatological methods, 2. Numerical modelling, and 3. Increasing the amount of observation points.

Thermal mapping is a road climatology based technique used in many countries to measure spatial temperature variation along the roads (Thornes, 1991). In the thermal mapping process, road surface temperature is measured with a vehicle attached instrument along the studied area. The measurements are done several times during night-time. The obtained temperature profiles can be used to adjust forecasted temperatures on the route. A drawback of the method is that a large amount of surveys is needed to cover different winter weather types well enough (Chapman and Thornes, 2006). In addition, thermal maps cannot provide information of the temporal development of the road surface temperature (Chapman and Thornes, 2006).

The second category contains methods that utilize information of local features in the road weather model. Geographical parameters, such as latitude, altitude, and land use can be used either directly in the model equations or as parametrizations (Chapman et al., 2001). Some of the parameters are simply obtained from geographic information systems, but others are not as easily available. Local horizon angle and sky view factor are parameters that can have considerable effect to the road surface temperature (Bogren et al., 2000; Chapman et al., 2001). Local horizon angle is the angle between the surface and the horizon at the road point. If the sun elevation angle is lower than the local horizon angle at the direction of the sun, the surface does not receive direct solar radiation. In practise local horizon angle can be determined as an average for certain range between angles, like from 0° to 5° and use the value for all directions in that range. The ranges should be small enough to timely determine the moment when the sun becomes obscured or visible, as it can affect greatly to the amount of radiation received. Sky view factor is defined as "The ratio of the radiation received (or emitted) by a planar surface to the radiation emitted (or received) by the entire sky hemisphere" (Oke et al., 2017). These parameters can be calculated for example from high resolution digital elevation dataset (Senkova et al., 2007) or determined from fish-eye imagery (Chapman and Thornes, 2004).

Increasing considerably the amount of observation points would make it possible to measure temperature variations along the road network in real-time. Observations would provide also more accurate starting state for the RWM. Installing a full equipped RWS would be rather expensive, but a more dense observation network could be achieved also with low cost sensors. These instruments would supplement measurements from RWSs and could be installed for example to the coldest road locations (Chapman and Bell, 2018). Even more cost effective method for obtaining more observations would be utilizing data from vehicles. Modern cars have multiple sensors and can provide weather and road condition related information.

Unfortunately, this data is rarely available. Nevertheless, additional devices can be installed into vehicles to gather information along the roads. Teconer RCM411 and RTS411 are this kind of instruments, measuring optically road condition, friction and road surface temperature (Haavasoja et al., 2012; Malmivuo, 2013). The data quality of the mobile observations should be assessed before utilizing them in road weather weather model initialization. Mobile observations are more exposed to disturbances than stationary measurements and for example heating from the vehicles can cause bias to the observations. In addition, ice and snow on the road can also affect to the optical surface temperature measurements as the measurement device does not detect the radiation from the actual road surface but that of the ice or snow layer.

Although observations improve the starting state of the RWM, the energy balance in the simulation might still not represent real conditions at the road point. This might be caused by local features not taken into account in the model or by the errors in the NWP model forecast used as input data. It is possible to adjust the energy balance in the model by utilizing road surface temperature observations. METRo (Model of the Environment and Temperature of Roads) model uses iterative method called coupling to tune the incoming radiation in the model (Crevier and Delage, 2001). The method has also been applied to the FMI RWM. The idea of the coupling method is simple, but several alternations can be done to the way how the adjustment coefficient for the radiation is determined. By fine-tuning the method, it is possible to reduce computation time and increase the surface temperature forecast accuracy.

Mobile surface temperature observations and their usability in the road weather forecasts has been studied from many aspects in this work. The first of the two research questions of this thesis is: "How to best use available surface temperature observations in the road weather model initialization?" Answering this question would enable more accurate road weather forecasts, which in turn would improve safety on roads as the maintenance operations could be planned well beforehand. Paper **I** included in this work studies which kind of effect the aforementioned coupling method has to the forecast accuracy. Paper **III** compares road surface temperature observations made with Teconer RTS411 to the RWS measurements. The study presents a road state (e.g. dry, wet or icy) dependent linear adjustment equation that can be used to adjust the RTS411 measurements. This equation is utilized in paper **IV**, which presents verification results of road weather forecasts utilizing observations made with RTS411.

The second research question of this thesis is "How differences in two road weather models' physics affect to the forecast accuracy?". Despite the fact that multiple RWMs have been developed during the past decades, there have not been many studies where two or more models would have been compared to each other. This kind of comparison would be beneficial for the model development as it could be analysed how differences in model physics affect to the road weather forecast. The

results can be utilized in the model development and can eventually lead to improved forecast accuracy. An opportunity for a comparison study occurred when the Royal Netherlands Meteorological Institute (KNMI) approached FMI with suggestion to compare the model results of the FMI RWM to the results of KNMI RWM. The KNMI RWM had been recently developed and there was a need to evaluate its performance to find out its suitability for forecasting road conditions in the Netherlands. Paper **II** of this thesis compares the verification results of the FMI RWM and the RWM developed in the KNMI.

2. Road weather model

2.1. Input and output

The FMI RWM is used to forecast road surface temperature and the amounts of water, ice, snow and deposit on the road (Figure 1). The model uses these values to determine also road condition, traffic index and friction. Road condition can have the following values: dry, damp, wet, wet snow, deposit, partly icy, icy and dry snow. Traffic index describes the overall traffic conditions and can be ”normal”, ”difficult” or ”very difficult”. Friction is calculated based on a statistical equation (Juga et al., 2012). The RWM requires 2 m temperature, relative humidity/dew point temperature, wind speed, incoming long and short wave radiation, precipitation and precipitation phase as input. If precipitation phase is not available it can be determined in the model.

The model run consists of two phases (Figure 2). In initialization phase, the atmospheric variables are obtained from observations. The short wave and long wave radiation are taken from a NWP model forecast due to the lack of measurements. The purpose of the initialization phase is to get a good starting state for the temperature profile in the ground. Its usual length in the FMI RWM is 2 days. If coupling method is used, there is an iterative radiation adjustment period at the end of the initialization phase. More information about this coupling phase is given in section 2.4. After the initialization there is a forecast phase, in which atmospheric values are taken from the NWP model forecast. The forecast phase does not start immediately at the NWP model

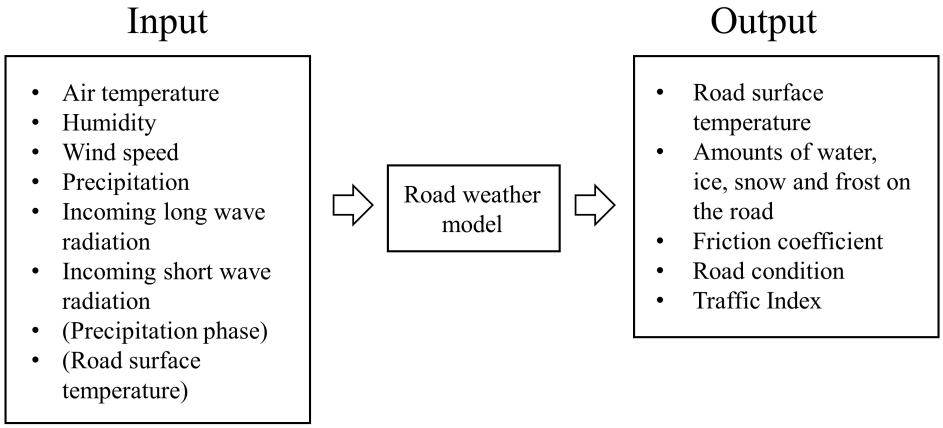

Figure 1: Road weather model input and output values. Precipitation phase and road surface temperature are optional. Input road surface temperature is used only in model initialization. Adapted from paper **IV**. ©American Meteorological Society. Used with permission.

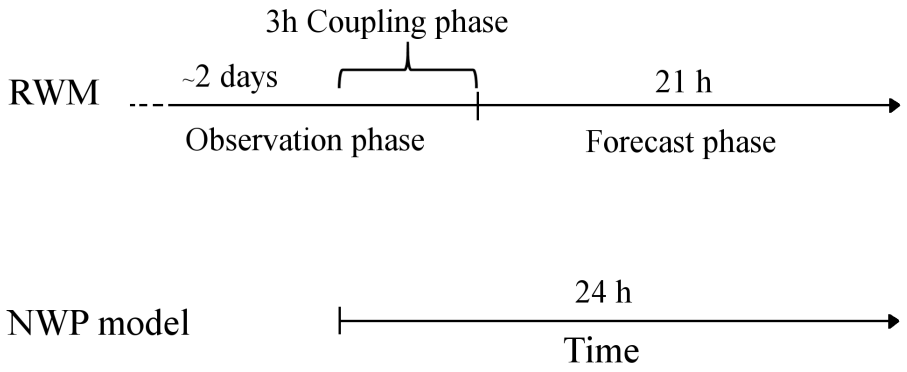

Figure 2: Phases in the road weather model and the relation of the model time line to a forecast from 3D numerical weather prediction (NWP) model. The duration of the model phases and the NWP forecast length may vary.

initialization time because the calculation of the NWP forecast takes considerable time during which new observations are made.

2.2. Surface temperature calculation

2.2.1. Heat transfer in the ground

The RWM is one-dimensional and consists of 16 layers. The two uppermost layers (0 - 1.5 cm and 1.5 - 4.75 cm) are asphalt layers and the rest have ground properties. The thickness of the layers increases with depth. The bottom of the lowest layer was set to 428 cm in the studies included in this thesis. Figure 3 illustrates the depth of the ground layers. The output surface temperature value is given as the average of the first two layers. The calculation of heat transfer between the layers is based on equation (Patankar, 1980):

$$\rho_g c_g \frac{\partial T(z,t)}{\partial t} = \frac{\partial}{\partial z} K \frac{\partial T(z,t)}{\partial z}, \tag{1}$$

where T is temperature, z is vertical distance in the ground, t is time, K is heat conductivity, ρ_g is density and c_g is specific heat capacity of the ground. Following equation is obtained when equation 1 is integrated over the volume of the layer and the time step and discretised (Patankar, 1980):

$$\rho_g c_g \frac{\Delta z}{\Delta t}(T_i^{j+1} - T_i^j) = f\Big[\frac{K_i(T_{i+1}^{j+1} - T_i^{j+1})}{z_{i+1} - z_i} - \frac{K_{i-1}(T_i^{j+1} - T_{i-1}^{j+1})}{z_i - z_{i-1}}\Big] + (1-f)\Big[\frac{K_i(T_{i+1}^{j} - T_i^{j})}{z_{i+1} - z_i} - \frac{K_{i-1}(T_i^{j} - T_{i-1}^{j})}{z_i - z_{i-1}}\Big], \quad (2)$$

where Δz is the layer thickness and Δt is the model time step. Index j refer to time and index i to ground layer. K_i refers to heat conductivity between layer $i+1$ and i. It

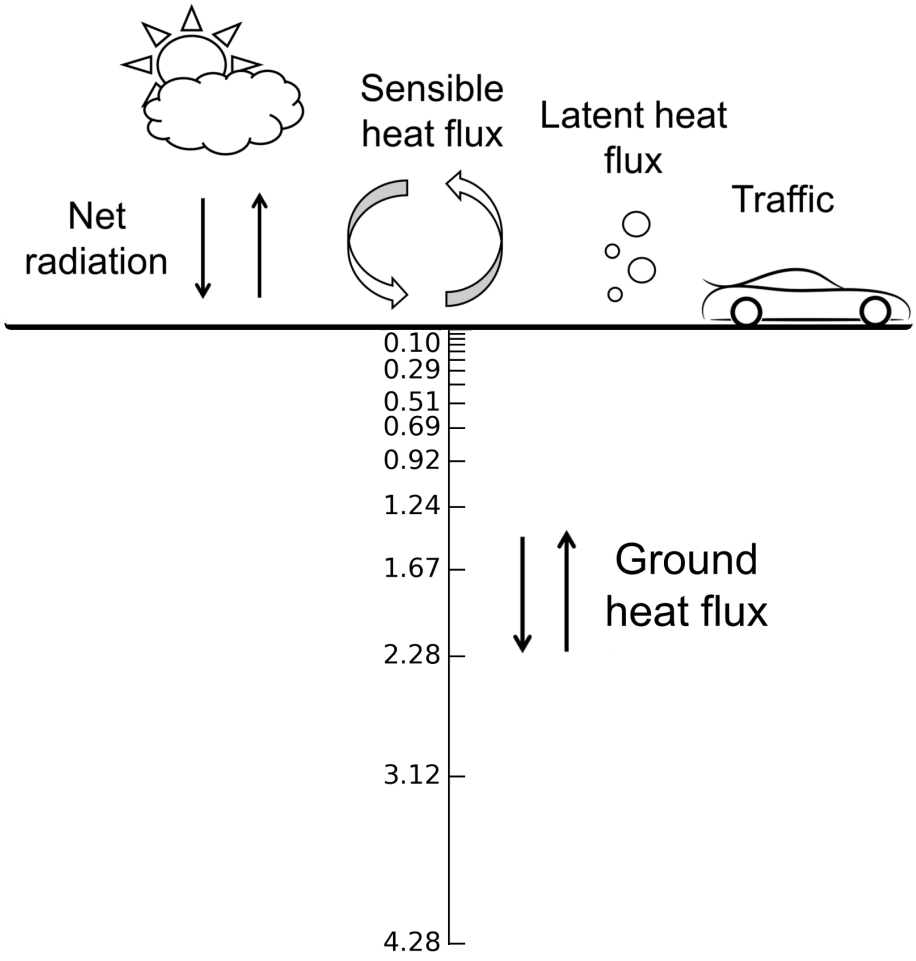

Figure 3: Illustration of energy fluxes in the road weather model and the depths of the ground layers (m). Adapted from paper **IV** ©American Meteorological Society

is assumed that the temperature varies with time according to formula:

$$\int_t^{t+\Delta t} T_i dt = [fT_i^1 + (1-f)T_i^0]\Delta t, \tag{3}$$

where f is a weighting factor between 0 and 1 (Patankar, 1980). The temperature of the ground layer changes with time, so the factor f determines how much weight is set for the beginning of the current time step and for the end of the time step. If f is set to 0, only the temperature at the beginning of the time step is used in the calculation. In that case the equation 2 has an explicit solution:

$$T_i^{j+1} = T_i + \frac{1}{\rho_g c_g \frac{z_{i+1}-z_{i-1}}{2\Delta t}}\left(K_i \frac{T_{i+1}^j - T_i^j}{z_{i+1}-z_i} - K_{i-1}\frac{T_i^j - T_{i-1}^j}{z_i - z_{i-1}}\right). \tag{4}$$

This solving method of equation 2 can be called forward difference explicit method (Campbell, 1985).

Some boundary layer conditions are required to solve the new temperature profile. The temperature for the lowest layer is calculated based on a climatological value:

$$T_b = T_c + A\sin(\omega J + \omega Di - \frac{z_b}{d}), \tag{5}$$

where T_c is the climatological temperature in the ground, A is the amplitude of the variation during the year, ω is variation frequency $(2\pi/365)$, J is the Julian day, Di indicates curve displacement, z_b is the depth of the bottom layer and d is damping depth. In the model $A = 0.6$, $Di = -170$ and $d = 2.7$ m. The upper boundary condition is defined as surface energy balance equation that is discussed in section 2.2.2.

Two different methods to calculate the new temperature profile in the ground have been used in the RWM. In the studies presented in papers **I** and **II**, the model used algorithm that aims to solve simultaneously the temperature for each layer. The equation 4 is arranged to form a function:

$$F = K_i \frac{T_{i+1}^j - T_i^j}{z_{i+1}-z_i} - K_{i-1}\frac{T_i^j - T_{i-1}^j}{z_i - z_{i-1}} - \rho_g c_g \frac{z_{i+1}-z_{i-1}}{2\Delta t}(T_i^{j+1} - T_i^j). \tag{6}$$

The Newton-Raphson method together with Thomas algorithm is used to iteratively solve the temperature profile (Fougstedt, 1992; Bristow et al., 1986; Campbell, 1985). New values for boundary layer conductance, latent heat flux and net radiation are calculated at each iteration step. This method was used in the model initialization phase with 5 min time step.

In the coupling phase and the forecast phase, the new temperature profile was calculated directly with equation 4 using 30 s time step. The reason for this change

Table 1: Surface and ground properties in the road weather model (Kangas et al., 2015)

Parameter	Value	Units
Specific heat, asphalt	919	J kg^{-1} K^{-1}
Heat conductivity, asphalt	0.5	W K^{-1} m^{-1}
Density, asphalt	2110	kg m^{-3}
Porosity, asphalt	0.1	
Specific heat, soil	813	J Kg^{-1} K^{-1}
Heat conductivity, soil	1.4	W K^{-1} m^{-1}
Density, soil	1600	kg m^{-3}
Porosity, soil	0.4	
Surface albedo, bare road	0.1	
Surface albedo, snow	0.6	
Surface albedo, ice	0.1-0.6	
Emissivity	0.95	
Absorption of long wave radiation	0.95	
Roughness length for momentum	0.4	m
Roughness length for heat	0.001	m
Temperature reference height	2	m
Wind reference heigth	10	m

was that the coupling method discussed in section 2.4. did not work well with the iterative solving method. In the study presented in paper **IV**, the model used the latter method and 30 s time step for the whole model run. The used values for model surface and ground properties are presented in Table 1.

After the research papers included in this thesis were published, a bug was discovered in the model code. Due to a misplaced parenthesis, when the next time step's temperature was calculated directly with equation 4, there was an error in the calculation of the term

$$\frac{1}{\rho_g c_g \frac{z_{i+1}-z_{i-1}}{2\Delta t}}. \tag{7}$$

Effectively, it caused the term to be calculated as:

$$\frac{\Delta t}{\rho_g c_g \left(z_{i+1} - \frac{z_{i-1}}{2}\right)}. \tag{8}$$

The effects of this error are discussed in section 4.

2.2.2. Surface energy balance

The energy balance at the surface is calculated with equation (Brutsaert, 1984):

$$G = I_{net} - H - \lambda E + Tr, \tag{9}$$

where G is heat flux into the ground, I_{net} is net radiation at the surface, H is sensible heat flux, λE is latent heat flux and Tr describes heating caused by traffic. Figure 3 shows a schematic representation of the model energy fluxes. The net radiation is calculated as:

$$I_{net} = (1 - \alpha_s)I_g + \epsilon_l I_L - \epsilon_s \sigma T_s^4, \tag{10}$$

where α_s is surface albedo, I_g is downwelling shortwave radiation, ϵ_l is absorption of longwave radiation, I_L is downwelling longwave radiation, ϵ_s is surface emissivity and σ is Stefan-Boltzmann constant.

2.2.3. Sensible heat flux

The calculation of sensible heat flux is based on boundary layer conductance (BLC) describing how easily heat is transferred in the boundary layer. The used equation for the H is:

$$H = BLC(T_s - T_a), \tag{11}$$

where T_a is 2m air temperature. BLC is calculated as (Kangas et al., 2015; Campbell, 1985, 1986):

$$BLC = \frac{c_a \rho_a k u_*}{\ln \frac{z_T - d + z_h}{z_h} + \Psi_h}, \tag{12}$$

where c_a is specific heat of air, ρ_a is air density, k is von Kármán's constant, u_* is friction velocity, z_T is temperature measurement height, d is zero-plane displacement, z_h is roughness length for heat and Ψ_h is semi-empirical stability correction function for heat. Friction velocity characterizes the shear in the boundary layer. In the RWM it is calculated as (Campbell, 1985, 1986):

$$u_* = \frac{ku}{\ln \frac{z_W - d + z_m}{z_m} + \Psi_m}, \tag{13}$$

where z_W is wind speed measurement height, z_m is roughness length for momentum and Ψ_m is semi-empirical stability correction function for momentum. The stability correction functions are used to correct boundary layer conductance for the buoyancy effects. If boundary layer is stable, mixing is smaller than in unstable conditions. Ψ_m and Ψ_h are calculated by using a factor that describes the relative importance of thermal

and mechanical turbulence in boundary layer transport (Campbell, 1985). This factor is calculated as

$$\zeta = -\frac{kz_T gH}{c_a \rho_a T_a {u_*}^3}, \tag{14}$$

where g is gravitational constant ($9.81\ m/s^2$). In stable conditions, ζ is positive and Ψ_m and Ψ_h are calculated as (Campbell, 1985, 1986):

$$\Psi_h = \Psi_m = 4.7\zeta. \tag{15}$$

In unstable conditions, ζ is negative and

$$\Psi_h = -2\ln\left(\frac{1 + \sqrt[2]{1 - 16\zeta}}{2}\right) \tag{16}$$

and

$$\Psi_m = 0.6\Psi_h. \tag{17}$$

Sensible heat flux must be solved iteratively because ζ is a function of H. This is done by first setting $\Psi_h = 0$ and $\Psi_m = 0$ and calculating the first guesses for BLC, H and ζ. The obtained value for ζ is used to calculate Ψ_h and Ψ_m. These values are used to calculate new values for BLC, H and ζ. The iteration is continued until the absolute difference of BLC values between subsequent iteration rounds is smaller than 0.001 or the maximum number of iterations is reached.

2.2.4. Latent heat flux

Latent heat flux in the model is calculated as (Calder, 1990):

$$\lambda E = \frac{\rho_m c_a}{\gamma} \frac{e_s - e_a}{r_o}, \tag{18}$$

where ρ_m is density of moist air, γ is psychrometric constant, e_s is vapour pressure of the surface, e_a is vapour pressure of the air and r_o is aerodynamic resistance.

According to Tourula ja Heikinheimo (1998), r_o can be calculated in unstable conditions as:

$$r_o = \frac{\left(\ln\left(\frac{z-d}{z_m}\right) + \Psi\right)\left(\ln\left(\frac{z-d}{z_h}\right) + \Psi\right)}{k^2 u}. \tag{19}$$

This equation is used in the road weather model with some modifications. In the model Ψ is replaced with Ψ_m when added to term $\ln(\frac{z-d}{z_m})$ and with Ψ_h when added to term $\ln(\frac{z-d}{z_h})$. In addition, zero plane displacement is assumed to be small, but surface roughness is added to the logarithm. Aerodynamic resistance is set to be 30 s/m in the model if wind speed is smaller than 1 m/s, because the equation gives rather

large values at small wind speeds. However, it was observed that this may cause abrupt changes in the aerodynamic resistance, as equation 2.2.4. can give values over 500 s/m at wind speed 1.1 m/s. In the study presented in paper **IV**, the limit was not dependent on wind speed, but the aerodynamic resistance had instead a maximum limit of 250 s/m.

2.3. Relaxation

Differences between observations and values obtained from a NWP model forecast can cause unnatural leap in the RWM input data. The resolution of a NWP model is usually from one to several kilometers and thus cannot take into account all local features. In addition, the NWP model forecast can be already several hours old when it is used in the RWM, which can increase the forecast error. Relaxation method can be used to smoother the transition from the initialization phase to the forecast phase. When the model starts the forecast phase, air temperature, wind speed and humidity values are adjusted according to an equation (Crevier and Delage, 2001):

$$X_i(t) = X_F(t) - (X_{FO} - X_O)e^{-\frac{t}{t_c}}, \tag{20}$$

where X_i is the adjusted value, X_F is the forecast value, X_{FO} is the forecasted value at the time of the last observation, X_O is the last observed value, t is the time elapsed from the start of the forecast phase and t_c is e-folding time (4 h). For example, the weight of the X_O is reduced to 20% after six hours. Theoretical example of relaxation is shown in Figure 4. Paper **I** discusses this method and its effects to the road surface temperature forecast in more detail.

2.4. Coupling

Paper **I** of this thesis concentrates on a method called coupling and its effects to the road weather forecast. The main idea of the coupling method is to adjust the radiation values so that the modeled road surface temperature fits to the latest observed value (Crevier and Delage, 2001). This is done iteratively by going trough a three-hour-period in the model several times until a fitting correction coefficient is found. This period is called coupling phase and it is conducted at the end of the initialization phase.

The correction coefficient is given either for incoming short wave (SW) or long wave (LW) radiation depending on which one of them has higher intensity at the start of the coupling phase. The RWM assumes open sky conditions and does not take into account the surrounding objects such as trees and buildings. In reality, the screening effect of such objects affect considerably to incoming SW radiation. In addition, LW radiation from the surrounding objects affect to the surface heat balance. The radiation

adjustment aims to correct the error caused by these local features. As the intensity of SW radiation is rather low during wintertime in Finland, the coefficient is usually given for incoming LW radiation during winter. The correction coefficient is used in the forecast phase so that it approaches 1.0 according to the equation:

$$C_F(t) = 1.0 + (C_R - 1.0)e^{-\frac{t}{t_c}}, \tag{21}$$

where C_F is the coefficient used for the radiation at the current time step. Term t is the time elapsed from the end of the coupling period and t_c is same as in equation 20.

The advantage of the coupling method is that it makes possible to use even a single road surface temperature observation efficiently in the model initialization. Consequently, it enables using infrequent mobile observations effectively in the road

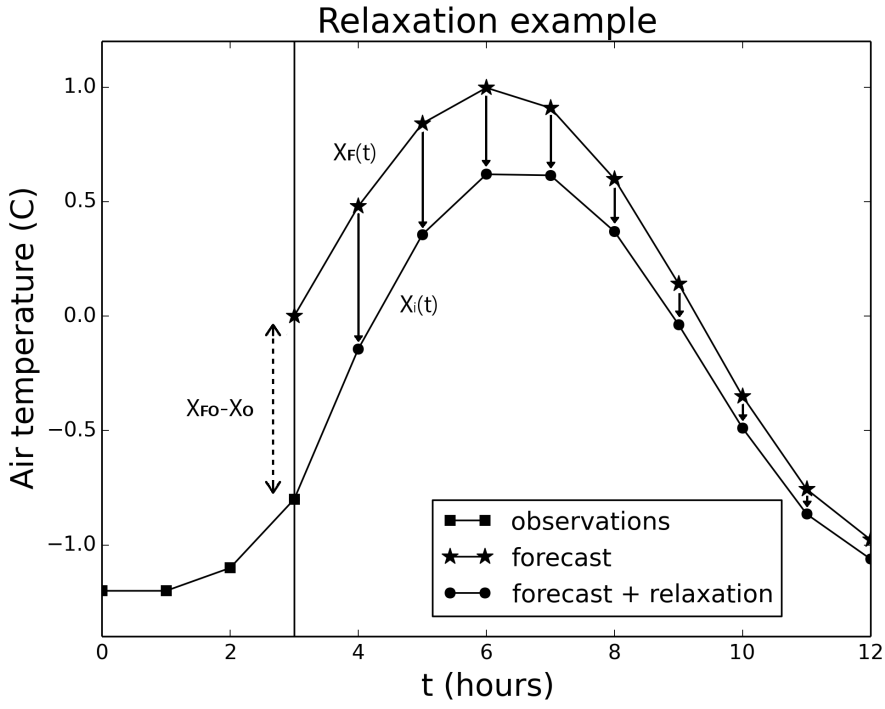

Figure 4: A theoretical example of relaxation. The curve with squares show the air temperature observations. The curve with stars shows the air temperature forecast ($X_F(t)$) and the vertical line the start of the forecast phase. Temperature difference between the forecast (X_{FO}) and the observation (X_O) at the time when last observation is made is illustrated with dashed line with arrows. The curve with dots shows the relaxed values obtained with equation 20 ($X_i(t)$). Adapted from paper **I**. ©2016 Royal Meteorological Society. Used with permission.

weather forecasts. However, there are some situations in which coupling method may cause problems. The correction coefficient may fit poorly to the forecast situation in the cases where the radiation is changing rapidly. This kind of sudden change may occur when the sun is behind a screening object during the coupling phase, but becomes visible at the beginning of the forecast phase. In this case the coupling coefficient will be fitted to the shadowed conditions and might cause too low SW values during the forecast phase. This can lead to considerably too low forecasted surface temperature values in the model.

It must be noted that although the correction coefficient is given for a radiation value, the actual error in the model might be caused by other factors. For example, if the forecasted wind speed is too low, the sensible heat flux can be too weak in the model. The coupling method can compensate also these other errors by increasing/ decreasing the radiation values.

2.5. Storage terms

The storage terms describe the amount of snow (S_s), ice (S_i), frost (S_f) and water (S_w) on road surface. The storages are increased/decreased by precipitation, melting, freezing, sublimation, desublimation and wearing. Kangas et al. (2015) describe the change in each storage term by conservation equations:

$$\frac{\partial S_w}{\partial t} = \Phi_{PR} - \Phi_E + \Phi_M - \Phi_F - W_w, \tag{22}$$

$$\frac{\partial S_i}{\partial t} = \Phi_F - \Phi_M + \Phi_S + K_{rf} W_s - W_i, \tag{23}$$

$$\frac{\partial S_{ii}}{\partial t} = \Phi_F - \Phi_M + \Phi_S + K_{rf} W_s - W_{ii}, \tag{24}$$

$$\frac{\partial S_s}{\partial t} = \Phi_{PS} - \Phi_M - \Phi_F - W_s, \tag{25}$$

$$\frac{\partial S_f}{\partial t} = \Phi_D - \Phi_M - W_f, \tag{26}$$

where Φ means source/sink terms and subscript PR refers to rain, PS to snowfall, E to evaporation, F to freezing, M to melting, S to sublimation and D to desublimation. W refers to wearing and blow off caused by traffic. Water wear includes also the water run-off. Factor K_{rf} represents the portion of the total snow wear caused by the snow packing to ice. The model has separate storage terms to describe the ice on the driving tracks and elsewhere on the road. The ice wearing term has higher value for the track storage than for the other ice storage.

Based on the storage terms, the model determines an output variable that characterizes road conditions (wet, icy, snowy etc.). More information of their determination is given by Kangas et al. (2015). In addition to the road condition, the RWM produces traffic index that describes overall driving conditions. It is determined based on the road condition, storage sizes, wind speed and precipitation intensity and type. The model calculates also road surface friction based on a statistical equation described by Juga et al. (2012).

3. Data

3.1. Observations

3.1.1. Road weather stations

There are around 400 road weather stations (RWS) in Finland maintained by Finnish Trasport Agency. Most of the RWSs are located in the southern part of the country where traffic density is considerably higher than in the north. The amount of measurement instruments in the RWSs varies, but typically they measure air temperature, humidity, precipitation, wind speed, surface temperature and road condition. Road surface temperature is measured in most stations with asphalt embedded Vaisala DRS511 (Vaisala, 2014). Some of the stations have more than one sensor to measure surface temperature on different lanes. DRS511 determines also road condition and the amount of water on the road. In addition to DRS511, many stations have optical Vaisala DSC111 (Vaisala, 2010). It is able to detect the presence of water, ice, slush and snow/frost on the road based on a spectroscopic measurement principle. The device determines also a classification for road condition (dry, wet etc.) and the level of grip on the road. Some road weather stations are equipped with Vaisala DST111 (Vaisala, 2017) instead of DRS511. DST111 measures road surface temperature by detecting the infrared radiation emitted by the road surface. Observations from the RWSs were utilized in the initialization and the verification of the RWM in the studies presented in papers **I** and **IV**. Paper **III** presents results where mobile observations discussed in section 3.1.2. were compared to RWS measurements.

Study presented in Paper **II** used observations from the Neatherland's RWSs. Approximately 300 RWS were included in the study. The stations measured surface temperature, conductivity, air temperature and dew point temperature. Soil temperature was also measured at some locations. Each station location had up to 12 road surface temperature sensors installed in the vicinity of the station.

3.1.2. Mobile measurements

Mobile measurements used in the studies presented in papers **III** and **IV** were done with Teconer RCM411 and RTS411 (Haavasoja et al., 2012; Teconer, 2016). They are both optical instruments that are installed on the vehicle. RCM411 measures road condition and friction coefficient based on the same measurement principle as Vaisala DSC411. RTS411 measures road surface temperature by detecting infrared radiation similarly as Vaisala DST111. The measurements are conducted once per second. The devices are connected to a nearby cell phone with Bluetooth connection. Information about location, direction and speed is acquired from the connected phone. The measurements were done between years 2014-2018 in Finland. The

cases where the device passed a road weather station were identified from the data to enable comparison between Teconer and RWS measurements. An average of RTS411 measurements done within 50 m range from the station were used in the comparison as a value for Teconer.

3.2. Atmospheric forecasts

Atmospheric forcing terms for the RWM forecast phase were obtained from forecasts generated by NWP model HARMONIE (Hirlam Aladin Research on Meso-scale Operational NWP in Europe)(Bengtsson et al., 2017). The model is nonhydrostatic and is based on Meteo-France AROME model (Seity et al., 2011). The used model configurations had grid-size of 2.5 km. Each of the studies presented in papers **I**, **II** and **IV**, had a different domain for the HARMONIE model. The domain covered Finland in paper **I**, The Netherlands in paper **II** and Scandinavia in paper **IV**. SURFEX model library is used to model surface processes within HARMONIE (Masson et al., 2013). The SURFEX library has different physical parametrisations depending whether the surface tile is classified land, town, sea or inland water. If the grid box contains more than one type of tile, weighed averages of the results for each tile are used as output values. The weights come from each tile's relative area in the grid box.

3.3. Kriging

Observations from road weather stations were used in the RWM initialization in the studies presented in papers **I** and **II**. In the study presented in paper **IV**, the idea was to investigate whether mobile observations improve the road surface temperature forecast in locations without RWSs. In this case, air temperature, humidity and wind speed values for the RWM initialization were obtained from a dataset of interpolated SYNOP observations. However, the forecasts were still done to the RWS points to enable forecast verification. The interpolation was done to 1 km grid with universal kriging method (Cressie, 1993; Aalto et al., 2013) by using sea and lake percentage and elevation as explanatory variables. Values for the simulation points were obtained with bilinear interpolation from the nearest grid points of the dataset.

FMI's operational road weather forecasts for points without RWS utilize road surface temperature values that are interpolated from road weather station observations. This interpolation is also done with universal kriging method, but the used grid size is 10 km. However, this data could not be used in the study, because the analysis included the observations from the RWS point where the forecast was made. To simulate a real case forecast scenario for a point without RWS, kriging analysis was performed again for each forecast point without using observations from that particular point. This interpolation was also done to the 10 km grid and the value for

the station point was obtained by bilinear interpolation from the nearest grid points. These values were used in the RWM initialization.

4. Summary of the results

4.1. The effect of coupling and relaxation to short range forecasts

Different model configurations utilizing coupling and/or relaxation were studied in paper **I**. Explanations of these methodologies are given in sections 2.4. and 2.3. The aim of the study was to find out how these methods affect to the road surface temperature forecast accuracy. Total 20 model configurations were tested on 20 road weather stations, including a reference configuration without coupling or relaxation. Some of the configurations used observed atmospheric values in the coupling phase and some used values obtained from a HARMONIE forecast. The length of the coupling period varied between the configurations. The model executions were conducted four times a day for time period 3 October 2013 to 13 January 2014. The simulations were not done in real time but as hindcasts afterwards, which was also the case in studies presented in papers **II** and **IV**. The road surface temperature forecasts were verified using observations from the RWS. The verification focused on the first forecast hour, as the effect of coupling and relaxation to the forecast is greatest at the beginning.

The coupling method clearly improved the verification results in most of the model configurations. Two model configuration types gave better results than the others. The first used forecasted atmospheric values in the coupling phase and did not include relaxation. The second used observed atmospheric values in the coupling phase and applied relaxation at the beginning of the forecast phase. Long wave and short wave radiation were obtained from the most recent HARMONIE forecast in the coupling phase with all model configurations as measurements were not available.

When values from a NWP are used in the coupling phase, also the forecast error in air temperature, wind speed and humidity affect to the energy balance and thus to the radiation correction coefficient. The radiation adjustment can thus fix forecast error also in these values. If observed atmospheric values are used in the coupling phase, the swift from the coupling phase to the actual forecast phase can cause abrupt changes in air temperature, wind speed and humidity values. The relaxation method smooths the transition from the initialization phase to the coupling phase. A widely used open source RWM METRo uses observed atmospheric values in the coupling phase with the coupling length of four hours (Crevier and Delage, 2001). In addition to air temperature, wind speed and humidity, relaxation is applied also to the surface pressure in the METRo model, which is not included in the FMI RWM input values. The forecasts verified in paper **IV** were also executed using this latter configuration but with coupling length of three hours. In the study presented in paper **II**, the FMI RWM applied the configuration using atmospheric values from HARMONIE in the

three hour coupling phase.

One important finding of the study was that the length of the coupling period does not affect much to the verification results. The reason for this was determined to be that the last hour in the coupling period affect considerably more to the radiation adjustment coefficient than the previous hours. Consequently, a shorter coupling period can be selected to save computing resources.

4.2. Comparison to the Netherlands RWM

Paper **II** represents results of the comparison between The Royal Netherlands Meteorological Institute's (KNMI) recently developed RWM and the FMI RWM. Both models were executed using the same input data for around 300 RWS points in the Netherlands. The executions were done four times a day for the test period 15 January - 28 February 2015. Overall, the KNMI's RWM produced slightly smaller error values when the forecast results were verified against road weather stations. The FMI RWM has been designed for Finnish wintertime conditions, whereas the physical properties of the KNMI model are optimized for the Netherlands. This partly explains the FMI model's slightly higher forecast error.

The verification focused on the forecasts starting at 03 UTC and 15 UTC. The utilized HARMONIE model executions were from 00 UTC and 12 UTC, respectively. The bias and RMSE values were calculated over all the stations for the full test period. The bias values have differences less than 0.1 K between models for the first 8-12 hours of these forecasts (Figure 5). The differences became larger with greater lead times. The models' bias values have clear daily cycle with lowest negative values at morning 09 local time (UTC +1 h, marked as LT henceforth) and in the evening at 19 LT. The largest positive values occur during the day around 13-16 LT and during the night around 05 LT. On average the FMI model is slightly colder than the KNMI model during midday, but there is great discrepancy between individual forecasts.

The RMSE differences between models are mostly around 0.1 K with the FMI model having greater values (Figure 6). The largest differences occur at midday when the RMSE values are the greatest. The difference is around 0.3 K at midday in the 03 UTC execution and around 0.5 K in the 15 UTC execution.

The RWMs produced surprisingly similar results regardless of the rather big differences in their physical properties. The thickness of the uppermost layer in the KNMI model is only 0.3 cm and the depth of it's lowest layer is 0.33 m. The corresponding values for the FMI RWM were 1.5 cm and 4.28 m. In addition, the FMI model's output surface temperature is the average of the first two layers, whereas in the KNMI model it is the temperature of the uppermost layer. The larger layer thickness in the FMI model would suggest that it is slower to react to the changes in the atmospheric conditions. However, when looking at the results calculated over all

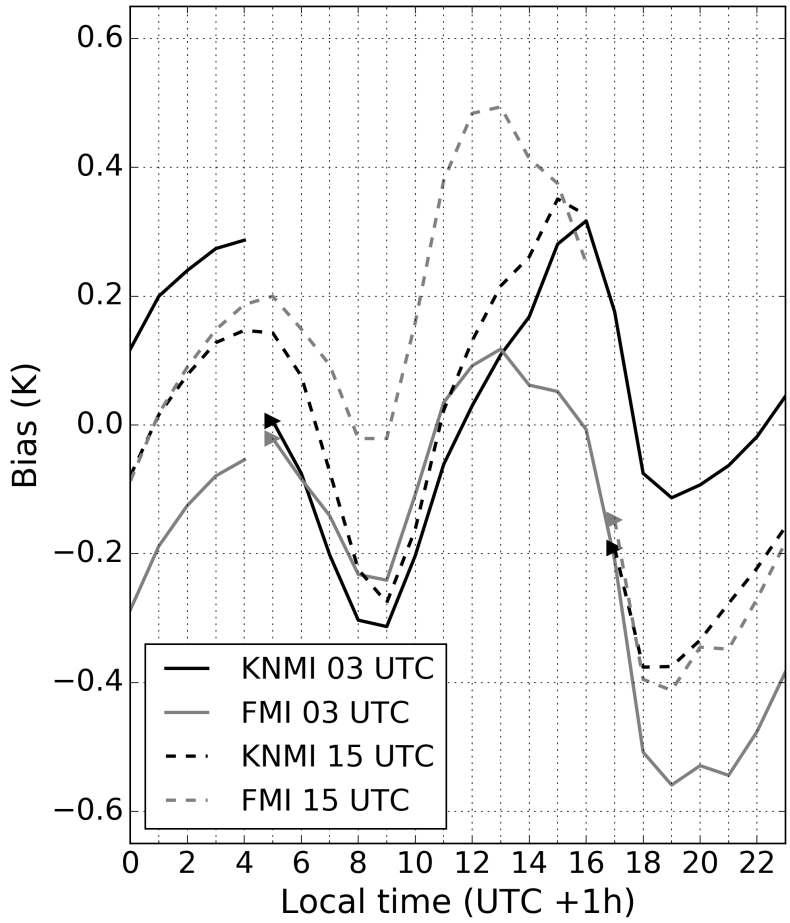

Figure 5: Bias of the road surface temperature forecasts as a function of local time. Continuous lines represent forecasts started at 03 UTC and dashed lines forecasts started at 15 UTC. Triangles show the time of the first forecast hour. The results of KNMI model are in black and the FMI model in grey. The figure is adapted from paper **II**. ©American Meteorological Society. Used with permission.

the stations, it seems that the FMI model cools faster during afternoon than the KNMI model and also warms up faster during morning. One reason for this behaviour is that the KNMI model has about two times bigger specific heat capacity value for asphalt than the FMI model (Table 1, Paper **II**). The asphalt density value is also considerably greater in the KNMI model. Consequently, more energy is needed to warm up the asphalt in the KNMI model. The FMI model has also significantly larger value for

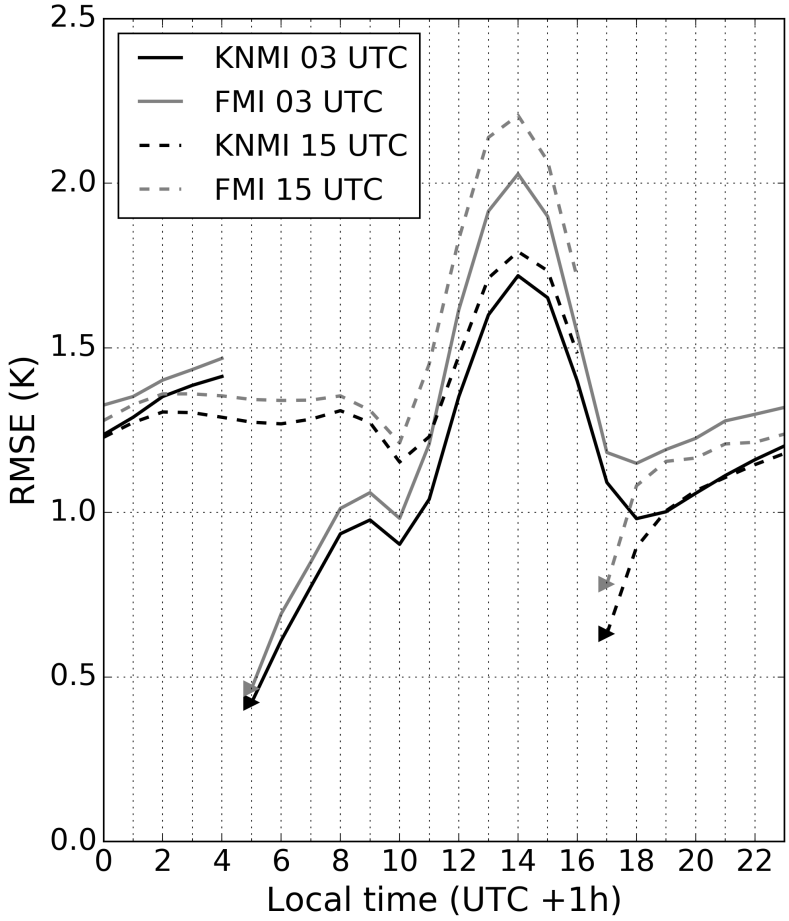

Figure 6: As 5 but for RMSE. The figure is adapted from paper **II**. ©American Meteorological Society. Used with permission.

roughness length of momentum. This causes stronger coupling between the road and the atmosphere and leads to greater absolute values of sensible heat flux. This supports the faster cooling during daytime in the FMI model because the air temperature is usually lower than the surface temperature.

The net effect of the physical properties and the used methodologies in the models lead to rather similar results regardless of the differences between models. However, the KNMI model's slightly better verification results show the importance to optimize the model physical properties to the climatological and physiological environments of the target area.

4.3. The usability of mobile observations in road weather forecasts

Mobile observations have great potential to improve forecasts in the areas without road weather stations. However, the data quality should be assessed before implementing new observations to the forecasting systems. In paper **III**, statistical analysis based on linear mixed effect models was used to find out possible differences between Teconer RTS411 and RWS surface temperature observations. The study used measurements from three winter periods between years 2014-2017 and included around 20 000 mobile-stationary observation pairs.

According to the results, Teconer observations were on average 0.62 °C higher than RWS observations in dry conditions and at 0 °C temperature. Different road conditions had significant effect to the surface temperature difference between the Teconer RTS411 and the Vaisala DRS511 instruments. When there is ice or snow on the road surface, the optical RTS411 does not detect the actual asphalt surface temperature but instead the long wave radiation from the ice or snow layer.

Above 4 °C, the RTS411 instruments gave increasingly lower surface temperature values than the RWSs. This was probably caused by the asphalt embedded DRS511 overheating in direct sunlight during daytime in august and spring. This is supported by the fact that similar behaviour was not detected when RTS111 was compared separately to the RWS measurements done with optical Vaisala DST111.

The data analysis showed also that the difference between the RTS111 and RWS measurements depends on the individual RTS111 device. The mean difference between the RTS111 and RWS observations were calculated separately for each device-station pair. At some station points, there were differences greater than 1.0 °C between values for different Teconer devices. This indicates that the installation and the vehicle have strong effect to the RTS111 measurements. For example, the infrared radiation emitted from warm engine bodies near the device can interfere the measurements if the radiation is reflected from the road surface to the sensor.

A linear adjustment equation was determined for the Teconer RTS111 observations to make them correspond better to the RWS measurements:

$$T_{s,Adj} = \beta_0 + \beta_1 \times T_{s,Mob}. \tag{27}$$

where $T_{s,Mob}$ is the mobile surface temperature, $T_{s,Adj}$ is the adjusted mobile surface temperature, β_0 is the level coefficient and β_1 is the slope coefficient. The values for the coefficients are dependent on the road condition and can be found from paper **III** in table 6.

The accuracy of the adjusted temperatures was evaluated with cross-validated mean squared error (MSE) calculated over the average of the MSEs of 10 unstratified

folds. The resulted MSE is 1.9 that is about 40 % smaller than the MSE of unmodified Teconer observations (3.3). Consequently, the RTS411 observations seem to be a potential data source for road weather modelling in areas with sparse RWS network. However, applying a statistical correction is suggested before implementing RTS411 observations to the forecasting system.

The effect of mobile observations to the forecast accuracy was tested in the study presented in paper **IV**. The forecasts were done for a period 12 October 2017 -30 April 2018. Mobile observations were utilized in the initialization whenever there was one available within three hours before the start of the forecast phase. This was done with the coupling method discussed in section 2.4. The model executions were conducted separately with unadjusted mobile surface temperature measurements and with three different adjusted values. The first adjusted value was obtained with equation 27 and the other two using equations determined from the same data as was used in paper **III**. The second adjustment equation is not dependent on the road state but includes individual intercept for each RTS411 device:

$$T_{s,Adj} = \beta_2 \times T_{s,Mob} + \beta_3. \tag{28}$$

Where β_2 is constant and β_3 depends on individual RTS411 device. The second equation includes both road condition based coefficients and device dependent intercept:

$$T_{s,Adj} = \beta_4 + \beta_5 \times T_{s,Mob} + \beta_6. \tag{29}$$

Where coefficients β_4 and β_5 are dependent on road condition and β_6 on RTS411 device. The values for coefficients are presented in paper **IV** in tables 1 and 2.

The model executions were done for RWS points to enable forecast verification. The verification results were compared to the results obtained from two control executions that did not use mobile observations. The first did not use surface temperature measurements at all while the second utilized interpolated values obtained with kriging analysis. More information about the kriging data is given in section 3.3. Operational road weather forecasts in the FMI utilize also similar interpolated values in points without RWS. Consequently, the second control execution represents best the currently used forecast method. However, the operational forecasts utilize time series formed from interpolated values, whereas the second control execution used only one value at corresponding time for the mobile observation. This was done to make the results more comparable. The verification scores were calculated separately for the ”winter” period (12 October 2017-17 February 2018) and the ”spring” period (18 February -30 April 2018). During springtime the verification results are much more dependent on the time of the day as the daily temperature variation is stronger.

Forecasting accurately near-zero temperatures is one of the most important aspects of the road weather forecasts. Consequently, the presented verification results include only cases where the RWS surface temperature was below 10 °C. This also reduced the error caused by RWS sensor heating too much during sunny days.

Model executions utilizing mobile observations gave clearly smaller RMSE values compared to the first control execution during the winter period (Figure 7). However, without the statistical corrections mobile observations increased considerably the warm bias in the model. The executions using adjusted mobile observations gave also somewhat smaller RMSE values than the execution using mobile observations without an adjustment. The second control execution using the interpolated surface temperature values gave rather similar RMSE values as the executions using the adjusted mobile observations during the winter period.

The daily temperature variation is large during springtime, causing difficulties to the RWM to forecast especially day-time maximum temperatures. The relative RMSE values between model executions were rather dependent on the forecast starting time. For example, the second control execution has rather small RMSE values for 09 UTC started forecasts. However, it gave the largest RMSE values for the 15 UTC started forecast. This behaviour is explained by the fact that the interpolated surface temperature values had tendency to be warmer than the mobile observations. These warmer temperatures compensated otherwise too cool model execution in 09 UTC started forecasts. However, the surface temperature cooled too slowly in the model during the evening in the 15 UTC started forecasts. In this case, the interpolated surface temperature values turned the already too warm model execution to be even warmer.

As a conclusion, assimilating the statistically corrected mobile observations to the RWM gives comparable results to assimilating interpolated values. However, the model executions were done within an area with high RWS density and small terrain variations. The mobile observations could improve the forecasts for example in northern Finland where the RWS network is sparse. Nevertheless, the results show that even a single road surface temperature observation can be used efficiently in the model initialization with coupling method.

The error in the model code mentioned in 2.2.1. effectively made the term multiplying the heat flux in equation 4 too small. This made the model layers change temperature more slowly than otherwise. In practice, fixing the error increases day time maximum temperatures and decreases night time minimum temperatures in the model during autumn and spring when the temperature has a clear daily cycle. However, correcting the error does not necessarily improve the forecast. In case of the study presented in paper **II**, the FMI had already warm bias during day time so the effect might have been negative to the model accuracy (Figure 5). Adjustments of the

model physical properties are currently under study to ensure the model performance after correcting the error.

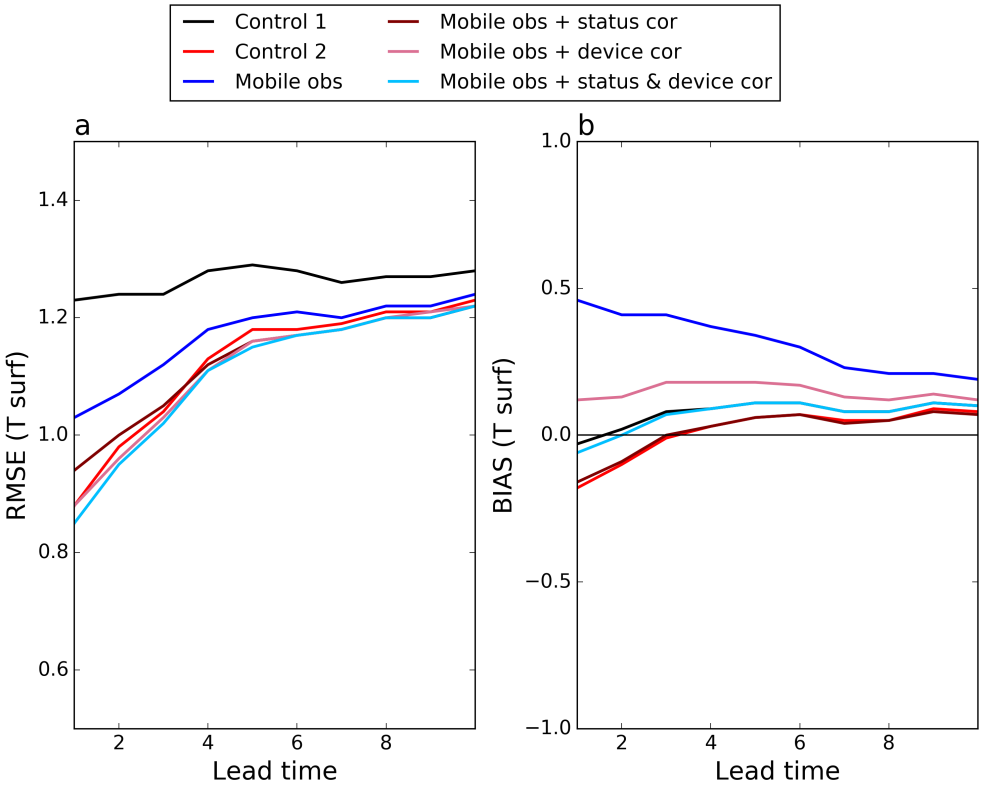

Figure 7: RMSE (panel a) and bias (panel b) values of the road surface temperature forecasts as a function of lead time. The results are calculated including all four forecast starting times for time period 12 October 2017- 17 February 2018. Only cases where RWS surface temperature was below 10 °C were included in the verification. Black line shows results for the first control execution, red line for the second control execution, blue line for execution using mobile observations, brown line for execution using mobile observations with status based correction, pink line for execution using mobile observations with device based correction and cyan line for execution using mobile observations with both status and device based correction. Adapted from paper **IV**. ©American Meteorological Society. Used with permission.

5. Conclusions and future perspectives

The main objective of this thesis has been to improve the accuracy of the road weather forecasts and thus improve the safety of roads during winter time. The presented forecast verification results have increased the understanding of the model behaviour and forecasting errors in different surroundings and forecast initialization times (Papers **I** , **II** and **IV**). One of the main research question of this thesis was: "How differences in two road weather models' physics affect to the forecast accuracy?" The studied models were the FMI road weather model and the Royal Netherlands Meteorological Institute's road weather model. It turned out that although the physical principles of the models are rather similar, there are some major differences in the ground physical properties between the models. The KNMI model has about two times bigger specific heat capacity value for asphalt than the FMI model. However, the net effect of the used methodologies and physical properties between the models lead to rather similar forecast results for road surface temperature.

The accuracy of road surface temperature forecasts of the KNMI RWM was slightly better than the accuracy of the FMI RWM. One reason for this was probably that the KNMI model was optimized for the roads in the Netherlands, whereas the FMI model has been designed to the Finnish conditions. The study highlights the importance of fine tuning the model physical properties so that the results fit to the observations done in the target area. Next time a similar change to compare performance of two RWMs occur, the author will consider fine tuning the model physical properties using historical data before executing the forecasts for the target time period.

The other main research question of this thesis was: "How to best use the available surface temperature observations in the road weather model initialization?". One can conclude that the coupling method (2.4.) is an effective way to use surface temperature observations and it improves the accuracy of short range forecasts. The method has been implemented to the operational RWM run at the FMI customer service. One of the advantages of the coupling method is that it enables the use of even a single available mobile observation in the model initialization. Paper **III** analyzes the quality of surface temperature measurements done with Teconer RTS411 instrument. The results show that surface conditions and the installation of the measurement device affect to the surface temperature measurements. Consequently, some statistical correction is recommended before using the measurements in the road weather model initialization (Paper **IV**). Consideration should be give also to how ensure the repeatability comparability of mobile measurements. This is a difficult task as the measurements are rarely done at exactly from the same spots of the road and the road temperature can vary also across the road cross profile. Moreover, the location information provided by the mobile phone connected to the instrument is not accurate enough to determine the measurement position exactly. It should be studied if there

is a method to detect the location more accurately.

As the availability of mobile observations increase in the future, it is important to take full advantage of them in the road weather forecasting. However, this study analyzed measurements only from one mobile surface temperature sensor type. As more measurements become available, it is important to assess the quality of the data from other sensor types. In the future, the validation of the measurements could happen also operationally on a cloud server by comparing multiple observation sources to each other. More studies are needed also to determine the best installation place for the sensor in different vehicles. Open places such as the back of the car are less disturbed by the radiation from warm bodies, but on the other hand are more exposed to dirt and snow.

In the future, model output statistics (MOS) (Glahn and Lowry, 1972) could be tested to adjust the model output to fit mobile observations instead of assimilating mobile observations to the RWM. However, this would require a lot of measurements from the same road stretches in different weather conditions to give reliable results. In addition, the statistical relationships between the parameters would require update always when changes are made to the RWM.

In addition to the efficient use of observations, attention should be given to better inclusion of local geography numerically in the RWM. Particularly important would be to determine values for sky view factor and local horizon angle on the main road network. Screening can affect greatly to road surface temperature especially during spring and autumn. These effects have already been included for example in the widely used METRo model (Kršmanc et al., 2014). They have been tentatively added also to the FMI RWM, but operational implementation requires still more testing and mapping the sky view factors and local horizon angles for the Finnish roads. The mapping could be done automatically for example by forming fish eye images from Google street view and determining sky and non-sky pixels by image processing (Middel et al., 2018). Microclimate effects such as cold air pooling in depressions affect also to the road surface temperature and should be better taken into account to get more accurate forecasts. However, as 3D features they cannot be modelled in an 1D RWM, so it would require statistical post processing either for the input obtained from 3D NWP model or the output of the RWM.

References

Aalto, J., P. Pirinen, J. Heikkinen, and A. Venäläinen, 2013: Spatial interpolation of monthly climate data for Finland: comparing the performance of kriging and generalized additive models. *Theoretical and Applied Climatology*, **112**(1-2), 99–111.

Bengtsson, L., U. Andrae, T. Aspelien, Y. Batrak, J. Calvo, W. de Rooy, E. Gleeson, B. Hansen-Sass, M. Homleid, M. Hortal, et al., 2017: The HARMONIE–AROME model configuration in the ALADIN–HIRLAM NWP system. *Monthly Weather Review*, **145**(5), 1919–1935.

Bogren, J., T. Gustavsson, M. Karlsson, and U. Postgård, 2000: The impact of screening on road surface temperature. *Meteorological Applications*, **7**(2), 97–104.

Bristow, K., G. Campbell, R. Papendick, and L. Elliott, 1986: Simulation of heat and moisture transfer through a surface residue-soil system. *Agricultural and Forest Meteorology*, **36**(3), 193–214.

Brutsaert, W., 1984: *Evaporation into the Atmosphere*. D. Reidel Publishing Company: Dordrecht, The Netherlands, 299 pp.

Calder, G., 1990: *Evaporation in the Uplands*. John Wiley and Sons, 148 pp.

Campbell, G., 1985: *Soil Physics with BASIC*. Elsevier, 150 pp.

Campbell, G., 1986: *An Introduction to Environmental Biophysics*. Springer-Verlag, 159 pp.

Chapman, L., and S. Bell, 2018: Network design considerations for a new generation of high resolution road weather information systems. *19th international road weather conference (SIRWEC), 29 May – 1 June, 2018*, Smolenice, Slovakia.

Chapman, L., and J. Thornes, 2004: Real-time sky-view factor calculation and approximation. *Journal of Atmospheric and Oceanic Technology*, **21**(5), 730–741.

Chapman, L., and J. Thornes, 2006: A geomatics-based road surface temperature prediction model. *Science of the Total Environment*, **360**(1-3), 68–80.

Chapman, L., and J. Thornes, 2011: What spatial resolution do we need for a route-based road weather decision support system? *Theoretical and applied climatology*, **104**(3-4), 551–559.

Chapman, L., J. Thornes, and A. Bradley, 2001: Modelling of road surface temperature from a geographical parameter database. part 2: Numerical. *Meteorological Applications*, **8**(4), 421–436.

Cressie, N., 1993: *Statistics for spatial data*. Jhon Wiley & Sons, New York, 900 pp.

Crevier, L., and Y. Delage, 2001: METRo: A new model for road-condition forecasting in Canada. *Journal of Applied Meteorology*, **40**(11), 2026–2037.

Fougstedt, B., 1992: *En studie över temperaturen i jordmånen (A study of the temperature in the soil)*. Master's thesis, University of Helsinki, in Swedish, 75 pp.

Fujimoto, A., A. Saida, and T. Fukuhara, 2012: A new approach to modeling vehicle-induced heat and its thermal effects on road surface temperature. *Journal of applied meteorology and climatology*, **51**(11), 1980–1993.

Glahn, H. R., and D. Lowry, 1972: The use of model output statistics (mos) in objective weather forecasting. *Journal of Applied Meteorology*, **11**(8), 1203–1211.

Haavasoja, T., J. Nylander, and P. Nylander, 2012: Experiences of mobile road condition monitoring. *16th international road weather conference (SIRWEC), 23 – 25 May, 2012*, Helsinki, Finland.

Jacobs, W., and W. Raatz, 1996: Forecasting road-surface temperatures for different site characteristics. *Meteorological Applications*, **3**(3), 243–256.

Juga, I., P. Nurmi, and M. Hippi, 2012: Statistical modelling of wintertime road surface friction. *Meteorological Applications*, **20**(3), 318–329.

Kangas, M., M. Heikinheimo, and M. Hippi, 2015: Roadsurf: a modelling system for predicting road weather and road surface conditions. *Meteorological applications*, **22**(3), 544–553.

Kršmanc, R., V. Tarjáni, R. Habrovský, and A. Šajn Slak, 2014: Upgraded METRo model within the METRoSTAT project. *17th international road weather conference (SIRWEC), 30 January – 1 February, 2014*, Andorra.

Malmivuo, M., 2013: Comparison study of optical friction and temperature meters 2013. Research reports of the Finnish Transport Agency 52/2013, Finnish Transport Agency.

Masson, V., P. L. Moigne, E. Martin, S. Faroux, A. Alias, R. Alkama, S. Belamari, A. Barbu, A. Boone, F. Bouyssel, et al., 2013: The SURFEXv7.2 land and ocean surface platform for coupled or offline simulation of earth surface variables and fluxes. *Geoscientific Model Development*, **6**, 929–960.

Middel, A., J. Lukasczyk, R. Maciejewski, M. Demuzere, and M. Roth, 2018: Sky view factor footprints for urban climate modeling. *Urban Climate*, **25**, 120–134.

Nurmi, V., A. Perrels, P. Nurmi, S. Michaelides, S. Athanasatos, and M. Papadakis, 2012: Economic value of weather forecasts on transportation–impacts of weather forecast quality developments to the economic effects of severe weather. *EWENT Deliverable Report*, **D5.2**, 92.

Oke, T. R., G. Mills, A. Christen, and V. J. A., 2017: *Urban Climates*. Cambridge University Press, 519 pp.

Päiviö, T., and O. Kärki, 2016: Talvikunnossapitoa Pohjoismaissa (Winter road maintenance in Nordic countries). *Tie & Liikenne*, **1/2016**, 18–21, in Finnish.

Patankar, S. V., 1980: *Numerical Heat Transfer and Fluid Flow*. McGraw-Hill, 197 pp.

Rayer, P., 1987: The meteorological office forecast road surface temperature model. *The Meteorological Magazine*, **116**, 180–191.

Seity, Y., P. Brousseau, S. Malardel, G. Hello, P. Bénard, F. Bouttier, C. Lac, and V. Masson, 2011: The AROME-France convective-scale operational model. *Monthly Weather Review*, **139**(3), 976–991.

Senkova, A., L. Rontu, and H. Savijärvi, 2007: Parametrization of orographic effects on surface radiation in hirlam. *Tellus A: Dynamic Meteorology and Oceanography*, **59**(3), 279–291.

Shao, J., P. Lister, G. Hart, and H. Pearson, 1996: Thermal mapping – reliability and repeatability. *6th International Road Weather Conference (SIRWEC), 17–19 April 1996*, Birmingham, UK.

Teconer, 2016: Road surface temperature sensor RTS411SA. Technical report, Teconer.

Thornes, J., 1991: Thermal mapping and road-weather information systems for highway engineers. *Highway meteorology*. E & FN Spon: London, 39–67.

Vaisala, 2010: Vaisala remote road surface state sensor DSC111. Technical report, Vaisala.

Vaisala, 2014: Vaisala DRS511 road sensor. Technical report, Vaisala.

Vaisala, 2017: Vaisala remote surface temperature sensor DST111. Technical report, Vaisala.

Yang, C., D. Yun, and J. Sung, 2012: Validation of a road surface temperature prediction model using real-time weather forecasts. *KSCE Journal of Civil Engineering*, **16**(7), 1289–1294.

I

Reprinted, with permission, from
Meteorologial Applications, 23, 503–513,
doi:10.1002/met.1574

METEOROLOGICAL APPLICATIONS
Meteorol. Appl. **23**: 503–513 (2016)
Published online in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/met.1574

RMetS
Royal Meteorological Society

Improving road weather model forecasts by adjusting the radiation input

Virve Karsisto,[a,b,*] Pertti Nurmi,[a] Markku Kangas,[a] Marjo Hippi,[a] Carl Fortelius,[a] Sami Niemelä[a] and Heikki Järvinen[b]

[a] *Finnish Meteorological Institute, Helsinki, Finland*
[b] *Department of Physics, University of Helsinki, Helsinki, Finland*

ABSTRACT: Considerable savings in winter road maintenance and accident costs can be achieved with accurate road weather forecasts. Forecasting road surface freezing time accurately enables the timely start of salting and thus ensures safety on roads. The optimal use of road weather observations is essential for the accuracy of short-range road condition forecasts. Road weather models incorporate radiation and other atmospheric variables from numerical weather prediction models. In this study, observations were used to correct the forecast radiation and thus improve road weather forecasts for a set of specific sites. Eighteen different configurations of this methodology were tested for 20 road weather stations in Finland during the autumn–winter period 3 October 2013 to 13 January 2014. This study shows that the coupling method has potential to significantly improve road surface temperature forecasts. Two model configurations in particular turned out to be better than the others giving almost equally good road surface temperature forecasts. It was found that the length of the adjustment period using the corrected radiation had only a slight effect on the results. The outcome of this study can be used to improve road weather models and thus achieve more accurate forecasts.

KEY WORDS road weather observations; road weather forecast; surface temperature; energy balance model; radiation correction

Received 30 June 2015; Revised 1 February 2016; Accepted 16 February 2016

1. Introduction

Road condition information is highly essential especially in the Nordic countries, where snow and ice cause difficulties for traffic every winter. Road weather forecasts that predict the condition of the roads are very helpful both for road users and the maintenance authorities, who need to make decisions on ploughing and salting actions. An accurate forecast can lead to considerable savings in road maintenance costs. For example, one needs much less salt to prevent a road from freezing than to melt the ice after freezing has occurred (Thornes, 1991). Timely maintenance operations also lead to better safety and more fluent traffic on roads. Road weather forecasts typically are based on 1D heat balance models, referred to as road weather models (RWM). They calculate the vertical temperature profile within the road layers and predict the road surface temperature (RST). The amounts of water, ice and snow on the road surface are also typically predicted. RWMs require atmospheric forecasts as input data, which are produced using a 3D numerical weather prediction (NWP) model. Air temperature, humidity, wind speed, precipitation and radiation parameters are interpolated from the NWP grid to the point at which the road weather forecast is produced. Usually, a RWM run has two separate phases: the initialization phase and the forecasting phase. In the initialization phase, the model uses observations as input data to obtain an accurate initial temperature profile in the ground. In the forecasting phase, this profile is forced by an atmospheric forecast in order to predict the evolution of the profile.

The transition from the initialization phase to the forecasting phase is very important for accurate short-range RWM forecasts. Even the first forecast hours often include error, because the atmospheric forecast is taken from a NWP model representing conditions integrated over large areas. The resolution of NWP models is often too coarse to take all local features into account, such as slopes and shadowing effects of forests. Therefore, the incoming radiation given by the model is calculated for open surroundings unless there are major topographical features in the area. However, these features need be taken into account in the RWMs, because they can lead to RST differences of up to 10 °C across a road network (Shao *et al.*, 1996; Bogren *et al.*, 2000). Even the smallest variations in atmospheric forcing can affect the freezing of roads at near-zero temperatures. Therefore, it is important to try to reduce the initial bias in the atmospheric forecast before applying it as the forcing for the RWM. In the case of radiation this is rather difficult, because there are no radiation observations available at road weather stations and the bias cannot be measured.

The local radiation amount can be modelled using sky-view factor, which defines the portion of sky that is visible (Oke, 1987; Chapman *et al.*, 2001a). Other more complicated methods, which give estimates of the time when the Sun is obscured (Chapman *et al.* 2001b; 2001c; Kršmanc *et al.*, 2014), have also been developed. However, determining the sky-view factors for Finnish roads would be time consuming and costly and the factors should be updated from time to time, for example due to forest logging. One option is to categorize road segments according to their surroundings and use the same sky-view factor for roads in the same category. The classified sky view should also be time-dependent, for example deciduous forest areas have a higher sky-view factor in winter than during summer because trees drop their leaves in autumn. Nevertheless, as long as the sky-view factor has not been estimated properly for the road network, other methods need to be used to correct the radiation forecast.

* Correspondence: V. Karsisto, Finnish Meteorological Institute, P. O. BOX 503, FI-00101, Helsinki, Finland. E-mail: virve.karsisto@fmi.fi

2. Review of forecast correction methods

There are several numerical methods to improve the forecast originating from an NWP model for a given location using available observations. One approach is to force the original predicted values towards the last observed ones. This relaxation method is explained further in Section 3.2. Relaxation is applied for air temperature, humidity, wind speed and pressure in the widely used Canadian open-source RWM called METRo (Model of the Environment and Temperature of Roads) (Crevier and Delage, 2001). The weight of forcing decreases exponentially from the observations to the forecast values over approximately a 6 h period.

The atmospheric forecast can also be corrected using a Kalman filter. This has been investigated by Homleid (1995), who used it in the LAM50 NWP model to reduce the bias in the surface temperature forecast. The principal idea is to use both observations and forecasts and combine them to make the best possible estimate of the target variable. This is done by using the linear minimum mean-squared error estimator. Recently, Habrovský and Tarjáni (2014) tested this method in METRo by using the Kalman filter to forecast the 2 m air temperature, which is commonly used as a forcing term to calculate the RST. This method was able to reduce the error in the air temperature forecasts according to their results; however, it was unable to follow rapid temperature variations.

Another statistical method used to correct the forecast is based on artificial neural networks (Shao, 1997). This algorithm is first trained with original meteorological variables and model errors. The neural networks will learn nonlinear dependencies between the model errors and the observed variables and after training it can correct errors in the forecasts. This method has been adapted successfully to the road ice nowcasting model 'Icebreak' (Shao *et al.*, 1993; Shao, 1997).

Sass (1997) presented a method in which a flux correction is applied to the heat balance equation so that the forecast surface temperature fits to the observed surface temperature. This correction is determined iteratively using different flux corrections and running the RWM only for a short period. The METRo model uses a similar kind of technique to correct the atmospheric forecast (Crevier and Delage, 2001). METRo includes a phase in between the initialization and forecasting periods called the coupling phase, in which both forecast values from the NWP model and observations are used (Crevier and Delage, 2001). This method has been applied and further developed in this study and is discussed in detail in this section. The calculation of the atmospheric forecast usually takes several hours during which new measurements are carried out. Therefore, there is usually a period at the start of the RWM calculation when both observations and forecasts are available. METRo uses the observed variables as forcing terms during this phase, except for radiation taken from the NWP forecast and RST calculated by the model. The model compares the calculated surface temperature with the observed one at the end of the coupling phase. If the difference is more than 0.1 K, the model sets a correction co-efficient for either incoming short wave (SW) or long wave (LW) radiation. The selection for which radiation variable the co-efficient is calculated is based on radiation intensities and will be detailed in Section 3.3. If the forecast temperature is too low, the co-efficient is set to greater than one so that the surface will receive more radiation than in the original run, thus becoming warmer. Conversely, if the forecast temperature is too high, the co-efficient is set to less than one. Thereafter, the model is returned to the start of the coupling period and will re-calculate the temperature evolution using this new co-efficient. This process is repeated by setting a new co-efficient for the selected radiation variable until the absolute surface temperature difference is below 0.1 K. After the coupling period, the calculated radiation co-efficient is used in the forecast phase so that exponentially it is reverted to 1.0 within approximately 6 h (Crevier and Delage, 2001).

The setting of the correction co-efficients for either incoming LW or SW radiation aims at incorporating the effects of local features at the forecast site. On the one hand, the radiation scheme in the atmospheric forecast does not take into account small local features, and in reality many road segments can receive less SW radiation than forecast due to shadowing objects such as trees and buildings. On the other hand, LW radiation from the surrounding objects can warm up the surface during the nights or in wintertime. The coupling phase can correct these model errors partly by fitting the correction co-efficient for one of the two incoming radiation variables. However, this method can cause problems when radiation changes rapidly between the coupling and the forecast period. For example, if cloudiness is much lower in the forecast than in the observations during the day, then the correction co-efficient will reduce radiation so that it fits the actual surface temperature during the day. If the sky becomes clear after the coupling period, the real surface temperature will rapidly increase, but the forecast surface temperature will remain low because of the radiation correction co-efficient.

Another problem of the coupling method is that although the last forecast surface temperature was within 0.1 K of the observed one, the forecast surface temperature generally differed more from the observations at other time steps during the coupling period. Therefore, it was investigated whether or not using a radiation correction co-efficient, which is based on time-dependent co-efficients for the coupling period, would improve the RST forecast. It is important that the ground temperature profile is as close to the real situation as possible after initialization. If the surface temperature in the model differs from the observed temperature during the coupling period, the whole temperature profile might not be close enough to the real one, even if the last surface temperature during the period is within 0.1 K of the observed one.

Different configurations of the coupling model were tested in this study. The aim was to determine the best way to calculate the co-efficient that would provide the best RST forecast. The data and the different coupling configurations are discussed in the next chapter.

3. Model and data

3.1. RoadSurf

The road weather model used in this study, RoadSurf, has been developed at the Finnish Meteorological Institute (FMI). It is 1D and uses the heat balance equation to calculate the road surface temperature. In addition, RoadSurf calculates the amount of water, snow, ice and deposit on the road. It also predicts road surface friction using a numerical-statistical equation (Juga *et al.*, 2013). As input, RoadSurf requires the forecasts of 2 m air temperature, relative humidity or dew point temperature, wind, precipitation and incoming SW and LW radiation. A more detailed description of the model can be found in Kangas *et al.* (2015).

3.2. Coupling configurations

RoadSurf was modified to evaluate different coupling configuration methods. Two reference configurations did not apply coupling at all and they transitioned directly from the initialization phase to the forecast phase. The length of the coupling

period in the other configurations varied from 1 to 4 h. The time of the first forecast hour was always kept unchanged; thus the coupling period started later in the configurations where it was shorter. Twelve configurations calculated the radiation correction co-efficient over the full coupling period, whereas six configurations calculated a separate co-efficient for each hour during the coupling period. The latter six configurations calculated the average of the co-efficients to be used in the forecast phase at the end of the coupling phase. Some configurations used only co-efficient averages of the last 2 or 3 h. Each of the separate co-efficients can be set either for LW or SW radiation. The final co-efficient is calculated as the average of the co-efficients set for the same radiation variable as the last hour's co-efficient.

It is possible to use either the forecast or the observed atmospheric variables as a forcing term during the coupling phase. The atmospheric variables included air temperature, dew point temperature, relative humidity, wind speed, precipitation and incoming SW and LW radiation. When using observations, only incoming LW and SW radiation are taken from an NWP forecast, because these variables are generally not available from observations. Thus, the calculated co-efficient aims to correct the error in the forecast radiation. Instead, when using forecast atmospheric values, the error in the forecast atmospheric forcing affects the co-efficient also. Out of the 18 models using coupling, 7 employed the forecast atmospheric values and the rest the observed values. In addition, 7 of the latter 11 model configurations used relaxation after the coupling phase meaning that the forecast air temperature, humidity and wind speed values were modified according to the bias of these variables at the time of the last observations. This method was implemented in RoadSurf following the METRo model approach (Crevier and Delage, 2001). Relaxation is performed using the following formula

$$X_i(t) = X_F(t) - \left(X_{FO} - X_O\right) e^{-\frac{t}{t_c}} \qquad (1)$$

where X_i is the forced value, X_F is the forecast value, X_{FO} is the forecast value at the time of the last observation, X_O is the last observed value, t is the time elapsed from the start of the forecast phase and t_c is a e-folding time (in this study $t_c = 4$ h). Consequently, the weight of the last observation reduces exponentially being 20% at 6 h. An example of the effect of relaxation is shown in Figure 1. In this case, the forecast is too warm at the time of the last observations, so relaxation lowers the forecast temperatures. One of the reference models, which did not use coupling, also used relaxation.

The main features of all 20 configurations are compiled in Table 1. It uses the following nomenclature for the different model configurations: Letters 'NC' means that the configuration does not use coupling, whereas 'C' means that coupling is used. The model configurations using forecast atmospheric values in the coupling phase include the letter 'F' and the model configurations using observed values the letter 'O'. The letter 'A' designates cases where the co-efficient is calculated using the averaging method explained above, and the letter 'R' indicates that relaxation is used. The number at the end of the codename refers to the length of the coupling period in hours, except for model configurations using an averaging method, where it means the number of co-efficients used to calculate the average.

3.3. Calculation of the radiation correction co-efficients

The coupling method used in METRo was implemented in RoadSurf with some changes. The radiation correction co-efficient can be calculated either for incoming LW or SW

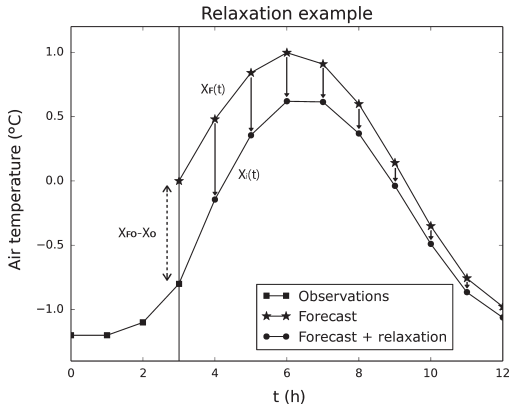

Figure 1. A theoretical example of the effect of relaxation on the original air temperature forecast (X_F (t), curve with stars). The forecast after relaxation is shown as curve with dots (X_i (t)).The vertical line shows the start of the forecast phase and the line with squares displays air temperature observations. Vertical dashed line with arrows represents the temperature difference between the forecast (X_{FO}) and the observation (X_O) at the time when last observation is made, which is used in Equation (1) to calculate X_i (t).

radiation. METRo applies SW radiation if it is above 100 W m^{-2} at the start of the forecast phase. It is furthermore required that SW radiation is above zero at least 3 h before the start of the forecast phase and 1 h after. Otherwise, the co-efficient is calculated for LW radiation. The method was changed in this implementation so that the co-efficient was calculated for the radiation variable which had the highest intensity at the start of the forecast phase. The Sun is above the horizon for only a few hours during the day in wintertime Finland and the incoming SW radiation intensities are quite low. Moreover, in northern Finland the Sun does not rise at all in mid-winter. Consequently, the intensity of LW radiation is usually much higher. The average daily maximum SW radiation intensity at the Kumpula measurement station in Helsinki was about 190 W m^{-2}, whereas the hourly average for LW radiation was about 300 W m^{-2} during the Octobers of 2008–2011. The corresponding values were 30 and 290 W m^{-2} in Decembers 2008–2011. The co-efficient was, therefore, set for LW radiation in most cases, giving a more stable correction, especially because the intensity of LW radiation varies much less than that of SW radiation. The used method has its drawbacks: if the forecast point is in shadow, usually it would be better to correct the low-intensity SW radiation, because then the correction would also disappear with sunset. If the correction is used for LW radiation in such a case, it would unnecessarily reduce the radiation after the error source has disappeared. However, it is not known in general whether the error is in the SW or LW radiation.

The value of 1.0 was used for the co-efficient (C_R) in the first iteration. If the forecast RST was not within 0.1 K from the observed RST, the C_R for the next iteration was determined using the procedure shown in the flow chart of Figure 2. It is based on the METRo source code (Canadian Meteorological Centre, 2013). The METRo procedure has been modified slightly in RoadSurf, as RoadSurf saves the first forecast RST and the corresponding C_R. RoadSurf replaces the saved values during each iteration, if the forecast RST is closer to the observed RST compared to the last saved value. The values are stored separately depending on whether the forecast RST is higher (T_{above}) or lower (T_{below}) than the observed RST. When both temperature

Table 1. Main characteristics of the various model versions.

Model version code	Use of coupling	Coupling time for calculating radiation co-efficients	Atmospheric data used in the coupling	Relaxation	Coupling period length (h)
1 NC	–	–	–	–	–
2 NCR	–	–	–	+	–
3 CF4	+	Over the full period	F	–	4
4 CF3	+	” ”	F	–	3
5 CF2	+	” ”	F	–	2
6 CF1	+	” ”	F	–	1
7 CO4	+	” ”	O	–	4
8 CO3	+	” ”	O	–	3
9 CO2	+	” ”	O	–	2
10 CO1	+	” ”	O	–	1
11 COR4	+	” ”	O	+	4
12 COR3	+	” ”	O	+	3
13 COR2	+	” ”	O	+	2
14 COR1	+	” ”	O	+	1
15 CFA4	+	” ”	F	–	4
16 CFA3	+	Average of each hour	F	–	3
17 CFA2	+	” ”	F	–	2
18 COAR4	+	” ”	O	+	4
19 COAR3	+	” ”	O	+	3
20 COAR2	+	” ”	O	+	2

The first column shows the version number and the second the codename. In the third column there is ‘+’ if the version used the coupling technique and ‘-’ if not. The fourth column indicates whether the radiation correction co-efficient is calculated over the full coupling period or whether coupling is done separately for each hour. In the latter option the model used the average of these co-efficients in the forecast phase. In the fifth column there is ‘F’ if forecast atmospheric variables were used in the coupling phase and ‘O’ if observed values were used. However, the radiation values were always taken from the forecast. The sixth column contains ‘+’ if the version used relaxation to smooth air temperature, humidity and wind speed values when moving from coupling phase to forecast phase and ‘-’ if relaxation was not used. The seventh column contains the length of the coupling period in hours.

values are stored, a new C_R is calculated assuming a linear dependence between RST and the radiation variable for which the co-efficient is used. This gives a good estimate for the next iteration, although in reality this dependence is not linear due to many connected processes. This procedure enables the fitting radiation co-efficient to be obtained with fewer iterations than using the original procedure from METRo.

When used in the forecast phase, the co-efficient gradually approaches 1.0 according to Equation (2):

$$C_F(t) = 1.0 + \left(C_R - 1.0\right) e^{-\frac{t}{t_c}} \quad (2)$$

where C_F is the co-efficient used for the radiation at the current time step and t and t_c are the terms as in Equation (1).

3.4. Data

The atmospheric forcing parameter values (2 m temperature, wind, humidity, precipitation, incoming LW and SW radiation) for RoadSurf were retrieved from the high-resolution NWP model, HARMONIE (Hirlam Aladin Research on Meso-scale Operational NWP in Europe). It is a nonhydrostatic model run operationally at FMI and is based on the Meteo-France AROME model (Seity *et al.*, 2011). HARMONIE (version cy36h14) has a horizontal grid size of 2.5 km and currently covers the whole of Scandinavia, but the coverage included only Finland at the time of the test period (3 October 2013 to 13 January 2014).

The atmospheric variables were interpolated from HARMONIE grid points to each of the 20 test stations located in different regions of Finland (Figure 3). These stations are equipped with the Vaisala ROSA road weather package. Fifteen stations measure RST with the Vaisala DST111 optical sensor (Vaisala, 2010), and the rest have Vaisala DRS511 sensors (Vaisala, 2001), which are installed in the road surface. The surroundings of the stations are described in Table 2, and the stations are grouped according to the openness to the Sun. Stations 1–5 are in open areas, Stations 6–16 are in areas where the sky is obscured in some directions but the southern horizon is relatively open, whereas Stations 17–20 are in areas where the surroundings clearly prevent SW radiation reaching to the surface during some part of the day. Table 2 also includes monthly average temperatures for each station during the test period. HARMONIE forecasts were available four times a day starting at 0000, 0600, 1200 and 1800 UTC. Each RWM configuration was run using atmospheric variables from a single forecast at a time so that a separate RWM forecast were made from all HARMONIE forecasts during the test period. The model runs were initialized with 48 h-long hourly observation time series obtained from road weather station measurements. At Station 10 precipitation and wind speed observations were missing and measurements of these variables were taken instead from a road weather station 8 km north from Station 10. Radiation was not measured at the stations and was taken from an older HARMONIE forecast, whose start time was close to the observation time. However, the radiation was not to calculate the RST unless the RST observations were missing.

Although mean temperatures were higher than usual in Finland in October, November and December 2013, the test period represents typical winter weather conditions with rapid temperature changes. However, there were also episodes with temperatures remaining nearly constant for weeks. On the one hand, the intensity of incoming SW radiation is very low in winter due to the northern location and clouds often reduce even further the fluxes reaching the surface. Consequently, daily temperature variations can be small. On the other hand, the air mass over Finland can change rapidly due to overpassing cyclones leading to quick changes in surface air temperature. Therefore, the behaviour of the model configurations was studied under different and representative weather conditions.

4. Results and discussion

4.1. Effect of radiation to the road surface temperature

Sensitivity test with the RWM were made at all stations for the December month to get a general view about how reducing or increasing the radiation affects to the RST. Reducing incoming LW radiation by 20 W m^{-2} over 1 h from the original value reduced RST by approximately 0.36 °C at the end of the hour when compared to the original model run. The test was repeated by decreasing the incoming LW radiation by 50 and 100 W m^{-2} and also increasing it by the same amounts. The results shown in Figure 4 indicate that the average dependency between LW radiation and RST is close to linear at least within this range of changes.

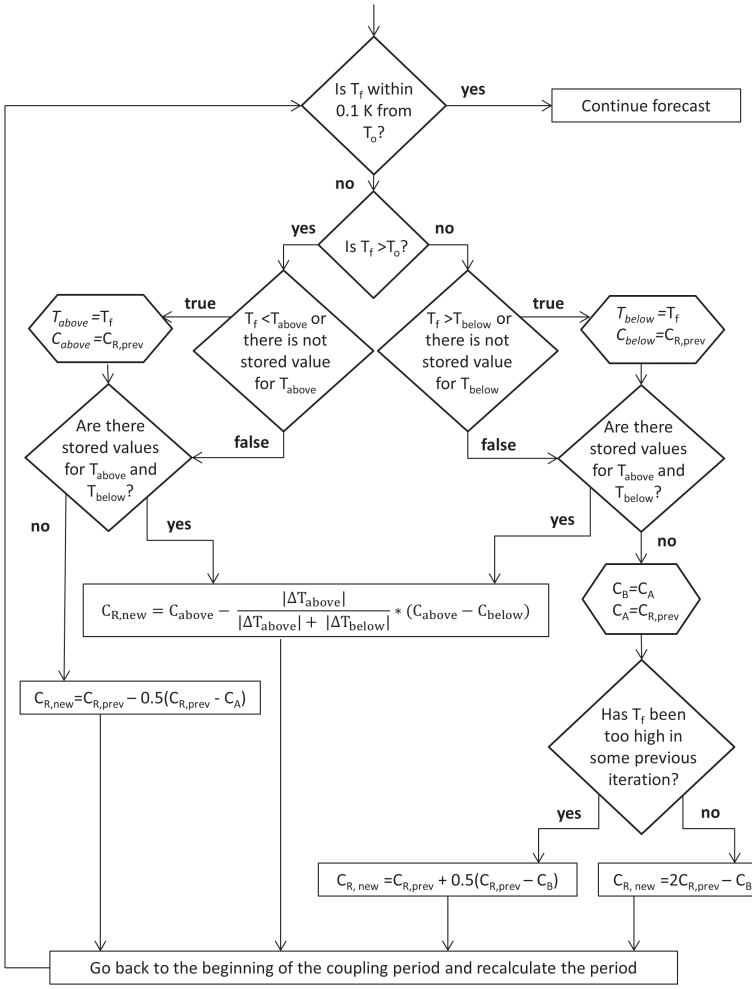

Figure 2. Flow chart of the determination of the radiation correction co-efficient. Symbols have the following meanings: T_f, forecast road surface temperature; T_o, observed road surface temperature; $C_{R,new}$, radiation correction co-efficient for the next iteration; $C_{R,prev}$, co-efficient from the previous iteration; C_A, co-efficient from the previous coupling iteration when T_f was too low. At the beginning $C_A = 0.0$; C_B, co-efficient from the second-to-last coupling iteration when T_f was too low. At the beginning $C_B = 0.0$. T_{above}, T_f from the iteration where the T_f was nearest to T_o and T_f was too high. T_{below}, T_f from the iteration where the T_f was nearest to T_o and T_f was too low. C_{above}, co-efficient which was used to obtain T_{above}; C_{below}, co-efficient, which was used to obtain T_{below}. $|\Delta T_{above}|$, absolute difference between T_{above} and T_o. $|\Delta T_{below}|$, absolute difference between T_{below} and the T_o.

4.2. Bias of RST forecasts

For road weather model runs that used data from HARMONIE forecasts at 0000 UTC, the first available forecast was at 0500 UTC. For the runs that used 0600, 1200 and 1800 UTC HARMONIE forecasts, the first valid forecast hours were 11, 17 and 23, respectively. Time is represented in UTC format, if not otherwise mentioned. Local time in Finland is UTC + 3 h until 26 October 2013 (EU summer time) and UTC + 2 h after that until the end of the test period. Bias of RST at the first forecast hour was calculated separately for each model configuration and for each station. The months of October and December were selected for detailed analysis because they represent typical Finnish autumn and winter weather conditions.

Figure 5 shows the bias (model-observation) for October and Figure 6 for December. The main observations are following:

(a) different model configurations usually yield biases of the same sign at the same station;
(b) at some stations, the different configurations have only small differences in bias values;
(c) configuration NC and CO1–CO4 have the largest absolute bias values;
(d) configuration CF1–CF4 and COR1–COR4 have the smallest absolute bias values in October, but in December COAR2–COAR4 give almost as small absolute bias as them;

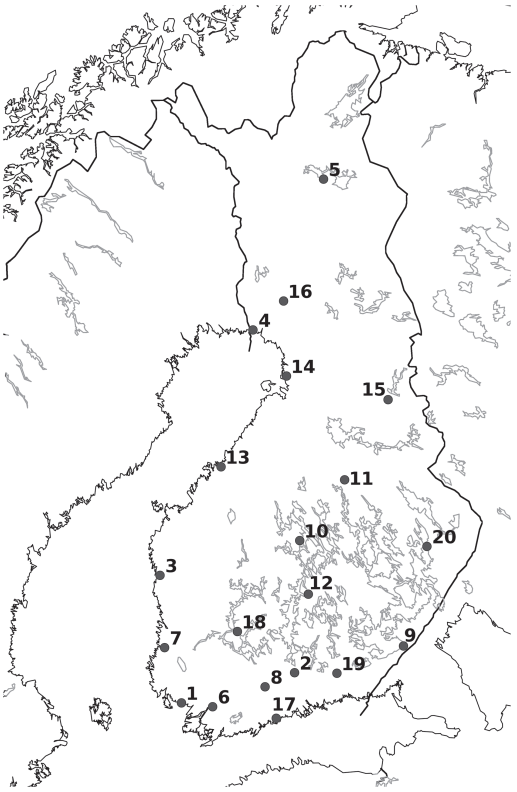

Figure 3. Locations of the road weather stations used as test site in this study. The numbers are the same as in Table 2.

(e) bias values between configurations that are otherwise the same but have different coupling lengths do not differ much, and
(f) there are no considerable differences between station groups 1–5 (open area), 6–16 (slightly obscuring surroundings) and 17–20 (obscuring surroundings) in the figures.

Point (a) indicates that whether the bias is negative or positive is very site-specific and that the different coupling configurations do not reverse it in most cases. Different configurations do not even affect the absolute value of the bias much at some stations, as noted in point (b). However, there are still clear differences between model configurations. According to point (c), almost all coupling methods reduce the bias notably when compared to the reference configuration. The only exceptions are the ones that use observed atmospheric values without relaxation (COi). It seems that without relaxation the forecast air temperature differs too much from the observations and causes large bias also to RST. In such a configuration, the error in the forecast air temperature did not affect the radiation correction co-efficient, because the observed air temperature was used to drive the RWM during the coupling period.

There are two model types, as stated in point (c), which gave better results than the others in October: coupling using the forecast atmospheric values (CFi, i = 1–4) and coupling using the observed atmospheric values with relaxation (CORi). Nevertheless, in December, the model configurations that use relaxation, coupling with observed atmospheric values and an average radiation co-efficient (COARi) yield as small bias values as CFi and CORi at some stations. Model configurations that use observed atmospheric values for the coupling and relaxation (CORi), seem to have slightly smaller bias than the model configurations using forecast atmospheric values (CFi). Furthermore, the length of the coupling period does not have any strong effect on the bias as noted in point (e). This is discussed more in Sections 4.3 and 4.4. The bias values have no clear differences between the station groups with different surroundings in October as mentioned in point (f). The errors in the NWP model and the other geographical factors, such as the elevation of the nearby surroundings, can also affect strongly the bias values. These errors and other geographical factors seem to have stronger effect on bias than open or obscured surroundings.

4.3. RMSE of RST forecasts

In addition to the bias, the root mean squared error (RMSE) of RST values was calculated for the first forecast hour and for each station and model configuration. The RMSE values of the first forecast hour RST for October 2013 are shown in Figure 7 and for December 2013 in Figure 8. The shading scale is logarithmic to highlight the differences between the model configurations. The main issues that can be seen from the figures are:

(a) in general, the error is larger in October that in December;
(b) the NC-model configurations has the largest RMSE values;
(c) configurations CF1–CF4 and COR1–COR4 give the smallest RMSE values during both months;
(d) configurations CO1–CO4 have larger RMSE values than COR1–COR4 during both months;
(e) configurations CFA2–CFA4 and COAR2–COAR4 have relatively large RMSE values compared to nonaveraged counterparts;
(f) RMSE values between the same model configurations but that have different coupling lengths do not differ much;
(g) Stations 17–20 (located in obscured area) have relatively large RMSE values in October but not in December, and
(h) there are no considerable differences between Stations 1–5 (open area) and 6–16 (slightly obscured surroundings) during both months.

Larger error in October than in December is due to lower daily temperature and radiation intensity variations in December. The RWM has often difficulties to predict the highest and lowest RST values even with the coupling method when the daily temperature variation is large, because the RST values are so sensitive to the total radiation. Moreover, RoadSurf is designed to work best in temperatures near zero because precise forecasts about very high and low temperatures are less important for winter road maintenance.

According to (h), all configurations using the coupling method have smaller RMSE values compared to the reference model without coupling and relaxation indicating clear RST forecast improvements by these methods. As can be determined from point (i), two of the same model configuration types that produced the smallest bias values also give the smallest RMSE values: CFi and CORi. Comparing these two types, the latter has distinctly smaller RMSE values at Station 4 in December, but there is no clear difference at other stations. Moreover, it was found that the larger RMSE values of configuration CF4 at Station 4 were caused mostly by the failure in coupling (i.e. the iteration did not converge to suitable correction co-efficient in some cases).

Table 2. Description of the surroundings of each test station with the mean observed air temperatures (°C) for each month during test period 3 October 2013 to 13 January 2014 (in October and January, only days 3–31 and 1–13 were used in the mean, respectively).

Station number	Co ordinates	Location and surrounding characteristics	Mean T (°C) October	Mean T (°C) November	Mean T (°C) December	Mean T (°C) January
1	60.4° N 22.3° E	Open area (SE–NW)	7.4	4.5	2.4	−0.6
2	60.9° N 25.6° E	Open area (SW–NE)	5.9	2.7	0.5	−1.9
3	62.3° N 21.5° E	Open area (SE–NW)	5.5	2.7	1.1	−2.8
4	65.8° N 24.3° E	Open, at the start of an overpass (E–W)	2.7	−2.2	−3.3	−5.7
5	68.0° N 26.9° E	Open, lake on both sides (SW–NE)	−0.0	−6.8	−8.6	−10.9
6	60.4° N 23.2° E	Few trees (SE–NW)	6.5	3.2	1.6	−1.4
7	61.3° N 21.7° E	Forest at the north, few trees in other directions (SE–NW)	6.2	3.4	1.3	−1.9
8	60.7° N 24.7° E	Empty lane between the road and forest on both sides (S–N)	5.9	2.5	0.5	−2.2
9	61.2° N 28.9° E	Empty lane between the road and forest on both sides (SW–NE)	6.0	2.5	−0.7	−2.7
10	62.8° N 25.8° E	Forest at the west, lake on the east (S–N)	4.1	0.5	−1.0	−3.6
11	63.7° N 27.3° E	Surrounded by few trees (SW–NE)	3.3	−0.5	−2.4	−5.0
12	62.0° N 26.1° E	Empty lane between the road and trees on both sides, lake nearby. (S–N)	4.8	1.5	−0.6	−3.4
13	63.9° N 23.3° E	Few trees (E–W)	5.3	1.7	0.2	−3.2
14	65.2° N 25.4° E	Empty lane between the road and forest in both sides (S–N)	3.3	−0.4	−2.3	−5.4
15	64.8° N 28.8° E	Trees on the north side of the road, open on other side (E–W)	1.9	−1.9	−4.2	−6.2
16	66.2° N 25.3° E	Trees on the west side, open on other side (S–N)	1.5	−3.9	−5.2	−6.6
17	60.3° N 25.1° E	Road with four lanes, forest on both sides (SW–NE)	6.9	3.8	1.6	−0.8
18	61.5° N 23.9° E	Forest on both sides (E–W)	5.6	2.4	0.4	−2.6
19	60.9° N 26.9° E	Empty lane between the road and forest on both sides (E–W)	5.5	2.4	−0.1	−2.3
20	62.6° N 29.8° E	Forest on both sides (S–N)	4.4	1.1	−2.0	−3.3

The orientation of the road is given in parenthesis, for example (SW–NE) means that the road orientation is southwest–northeast.

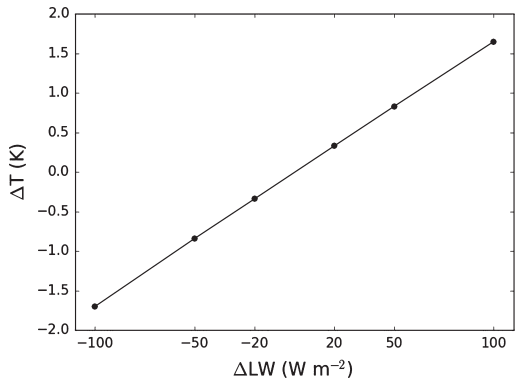

Figure 4. Average effect of changing the intensity of the long wave radiation (x-axis) for a 1 h period to the road surface temperature (y-axis) at the end of the hour. The test runs were performed for December and include all 20 road weather stations.

Point (j) implies that the relaxation procedure remarkably reduces RMSE values. The model configurations that use observed atmospheric values without relaxation (COi) give larger RMSE values than configurations with relaxation (CORi) at most stations. Point (k) reflects that the model configurations calculating the radiation correction co-efficient separately for each coupling hour and then using their average in the forecast phase give quite large RMSE values, although the RST values obtained during the coupling period fit well with the observed ones. The average co-efficient clearly does not fit as well for the first forecast hour as the co-efficient calculated over the whole coupling period. On average, each hour within the coupling period has the same weight, but when the co-efficient is calculated over the full period, the latest hour is the most significant as will be shown in the next section. It seems that when the last coupling hour has the strongest weight in the correction co-efficient, it fits the first forecast hour better than when each hour is weighted equally. One possible reason is that the heat balance conditions of the last coupling hour match better the next hour's conditions than those of the previous hours. Point (l) also indicates that the last coupling hour has the greatest weight when the co-efficient is calculated over the full period, because the length of the coupling period does not seem to affect much the RMSE values as it did not affect the bias either. The effect of the coupling length is discussed further in Section 4.4.

Contrary to the bias values the RMSE values have some difference between the station groups with different surroundings in October as mentioned in point (m). Stations 17–20, where the Sun was the most obscured during some part of the day, have larger RMSE values than the other stations on average. This can be due to the changes in the radiation caused by the shadowing objects, which prevent direct radiation from reaching to the surface during part of the day. When radiation changes after the coupling phase, the correction co-efficient does not fit the situation anymore and causes larger errors. However, as point (n) indicates, there is no clear difference between stations with open (Stations 1–5) and slightly obscuring surroundings (Stations 6–16). In December, there are no significant differences between RMSE values at different station groups due to the low intensity of the SW radiation.

This and the previous sections have focused on the verification of the first forecast hour, because it best illustrates how the transition from the initialization phase to the coupling phase has succeeded. The differences in RMSE values between the model configurations were smaller in the later forecast hours. At individual stations the relative quality between configurations changed as the forecast advanced, but as a whole there was no significant variation. It must be kept in mind that although coupling in most cases reduces the forecast error, it can in some cases cause larger errors than running the model without

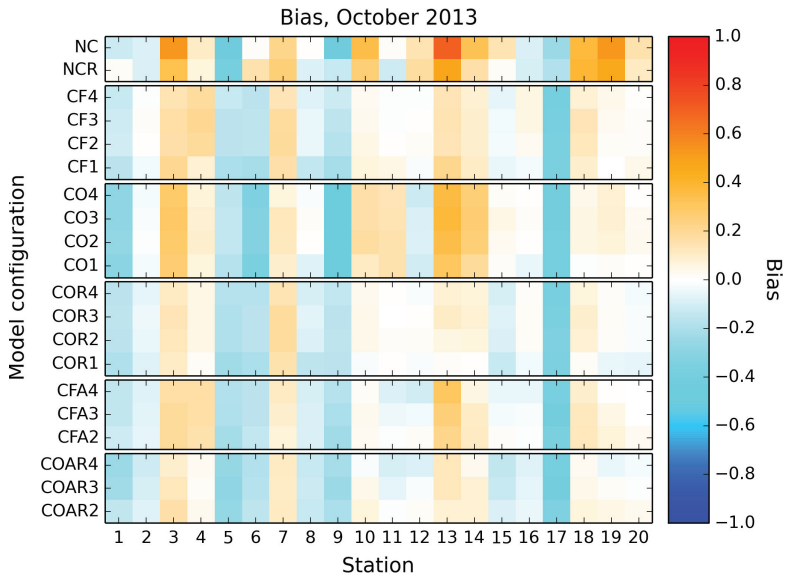

Figure 5. Bias of the road surface temperature (in K) for the first forecast hour during October 2013. The station indices on the *x*-axis refer to Table 2. The model configurations along the *y*-axis are described in Table 1. The configurations are divided to panels so that configurations that differ only in the length of the coupling period are in the same panel, except in the uppermost panel, that includes the configurations that did not use coupling. Positive bias values are in orange/red and negative in blue.

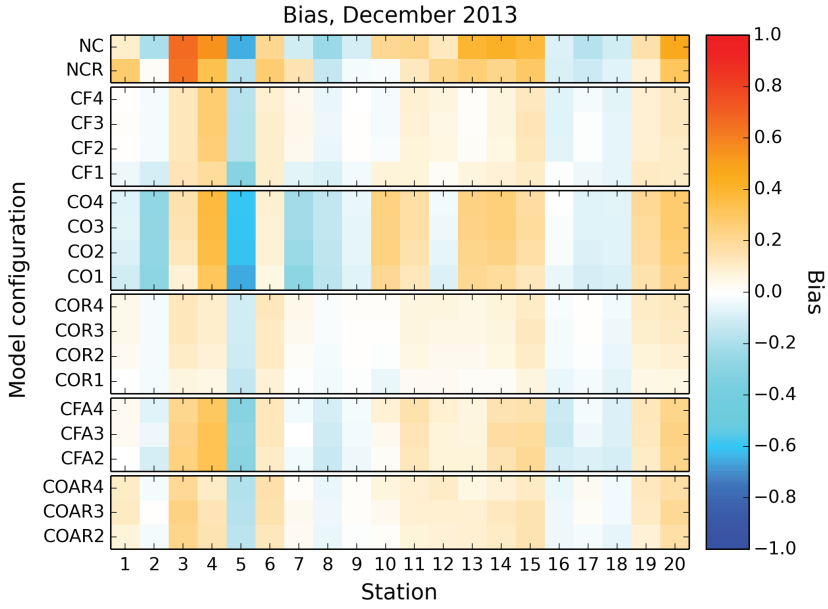

Figure 6. Same as Figure 5 but for December.

coupling. This can happen especially when radiation changes rapidly after the coupling phase.

4.4. Correction co-efficients

It seems that the length of the coupling period does not affect greatly the results based on the bias and RMSE values. This is an important result, because computer processing time can be reduced if the coupling period is not unnecessarily long: it is more beneficial to use a shorter coupling period if the results are equally good. The RMSE results show that a coupling duration of 2 h would be the most cost-effective choice. The obtained correction co-efficients were studied in more detail in order to get more information about the effect of different coupling lengths.

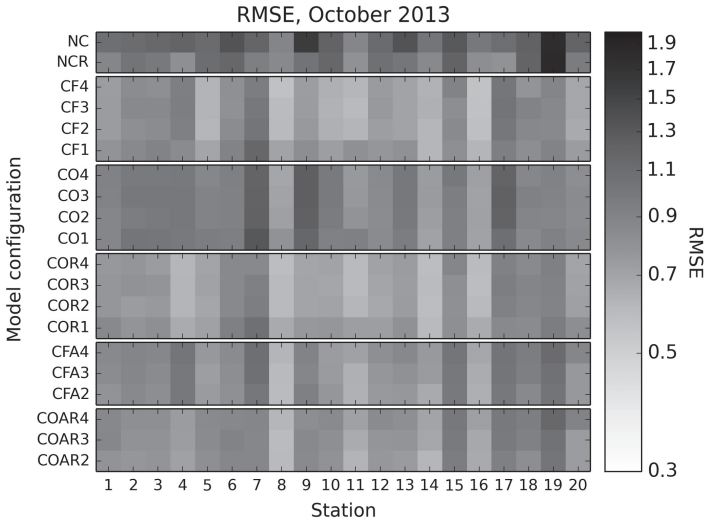

Figure 7. RMSE of the road surface temperature (in K) for the first forecast hour during October 2013. The station indices on the *x*-axis refer to Table 2. The model configurations along the *y*-axis are described in Table 1. The configurations are divided to panels so that configurations that differ only in the length of the coupling period are in the same panel, except in the uppermost panel, that includes the configurations that did not use coupling. The shading scale is logarithmic.

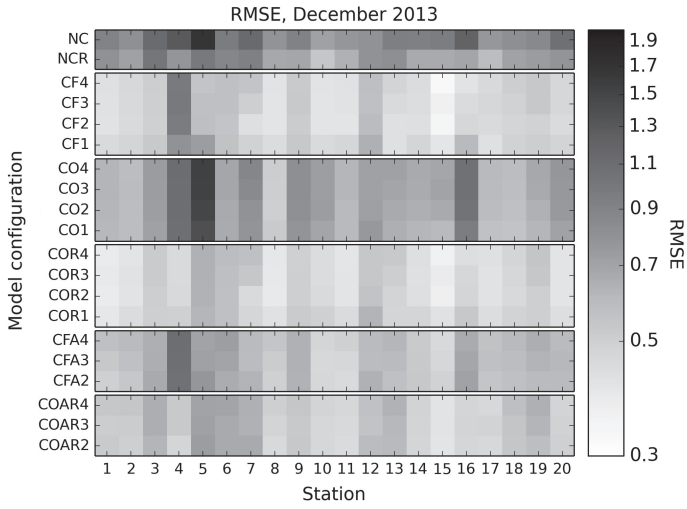

Figure 8. Same as Figure 7 but for December.

Figure 9 shows the co-efficients calculated using different coupling lengths with CORi (i = 1, 4) model configurations for all stations and for the whole test period (a–c). In addition panel (d) compares co-efficients obtained from the model configurations CF4 and COR4, which used different atmospheric variables during the coupling period. The figure contains only cases when the co-efficient was used for LW radiation, which was valid in 98% of all forecasts. It can be seen clearly that the more the length of the coupling period differs, the less the co-efficients agree. The co-efficients obtained from COR4 and COR3 are quite close to the diagonal line, but the co-efficients from COR4 and COR1 have larger differences. However, it can be seen that although the co-efficients differ quite much between COR1 and COR4 models, they are still more linked to each other than the co-efficients from model configurations using different datasets (Figure 7(d)). It can be concluded from the plots and average absolute differences between the co-efficients, that changing the length of the coupling period affects the co-efficients, but the effect is not large when the difference is only 1 or 2 h. When the difference is 3 h, the effect on the co-efficients starts to be more significant.

The model configuration COR4 was run with some modifications to test the significance of the coupling period length on the correction co-efficient. First, the model was changed so that it increases incoming LW radiation by 20% throughout the first hour in the coupling period. The model was then run for the

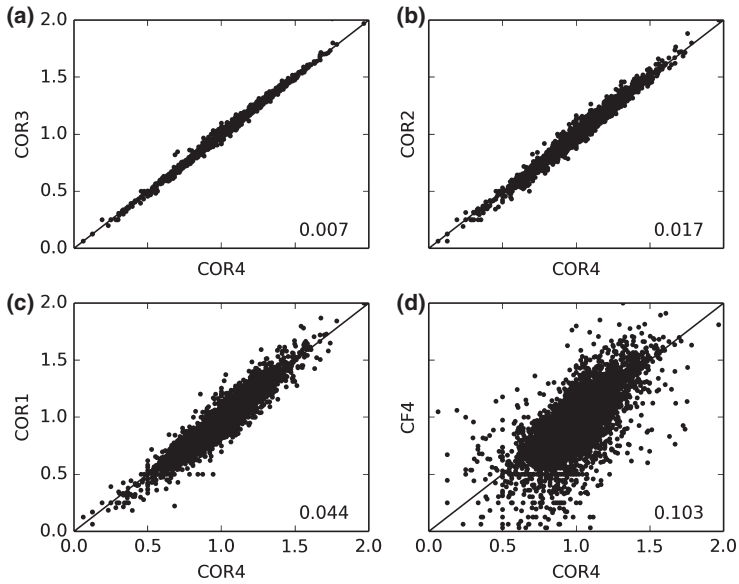

Figure 9. Comparison of the radiation correction co-efficients for long wave radiation obtained from different model configurations. The *x*-axis in each panel represents the co-efficients obtained with a 4-h-long coupling period (COR4). The *y*-axis represents the co-efficient obtained from (a) 3 h, (b) 2 h and (c) 1 h long coupling period. Panel (d) compares the co-efficients for configuration CF4 (*y*-axis) to COR4. The value displayed in at bottom right corner of each panel is a value showing the average absolute difference between the co-efficients in the panel. The Figure contains co-efficients from runs at all 20 stations and for the whole test period except for a few offset values. The black lines represent perfect agreement between co-efficients of both models.

whole test period and for all stations. The same test was performed three times by increasing incoming LW radiation during the coupling period: (1) throughout the second hour, (2) throughout the third hour and (3) throughout the fourth hour. When the first coupling hour radiation was changed, the mean change in the correction co-efficient was −0.010. The mean changes in the co-efficients when the second, third and fourth coupling hour LW radiation was increased were −0.019, −0.038 and −0.113, respectively. According to the results, the effect of increasing the last coupling hour radiation was threefold compared to the test where the second-to-last (third coupling hour) radiation was increased. This fact indicates that the last coupling hour has much more significance than the previous ones. Changing the last hour radiation affects directly the co-efficient, but if radiation only from previous hours is changed, the effect has already faded away when the model has adjusted to the situation afterwards.

5. Conclusions and outlook

This study has shown clearly that the coupling method has the potential to significantly improve road surface temperature forecasts. There are many ways to calculate the radiation co-efficient, but two model types gave the most accurate results with the set of road weather stations being used: configurations using observed air temperature, humidity, wind speed and precipitation values in the coupling phase and relaxation in the forecast phase, and configurations using the forecast values for these variables in the coupling phase. In both of these best performing configurations types the radiation co-efficient was calculated over the full coupling period. The former method is used in the METRo model (Crevier and Delage, 2001). One of the most important findings of this study is that the length of the coupling period does not seem to affect model accuracy considerably. This confirms that the last hour of the coupling phase dominates when determining the radiation correction co-efficients. This is probably the reason why the method that calculates a separate co-efficient for each hour in the coupling period and then uses their average did not give as good results. Another benefit from this study is that the numerical method to calculate the radiation co-efficient was improved to get the right co-efficient with less iteration.

Yet it must be kept in mind that these results were obtained from a limited set of road weather stations in Finland and from just one winter. It is possible that the model configurations would behave differently under different circumstances, for example by using different set of stations or time period. The performance of the coupling method also depends greatly on the time of day because of the changing radiation. Outputs were calculated only for four imposed hours of the day in this study, so the results might be different with different forecast settings. One restriction of the coupling method is that it requires precise RST observations. In many regions, road weather stations are far from each other and RST measurements are sparse. Consequently, the coupling method would not give the best possible results on road stretches between stations.

The advantage of the coupling method is that it does not require information about the surroundings of the road and is thus easy to implement to the RWM. However, if the shadowing effects could be taken into account, the model would become more realistic. The model performance would probably increase if the coupling could be used together with modelled screening effects, because then the coupling could correct the errors related to actual weather, like cloudiness, rather than the errors caused by shadows on the road. The screening effects should be taken into

account when developing the model in the future. Another essential component for the development of road weather forecasts is the continuously increasing accuracy of NWP models with higher resolution and better parametrizations of local processes. Consequently, this should lead to better road weather forecasts as well. The radiation terms forecast by NWP models are crucial for the coupling method. Therefore, as the NWP models develop (e.g. cloud–aerosol–radiation interaction), the RWM parametrizations and the coupling method must be further tested to ensure best model performance.

Acknowledgements

The authors are grateful to Professor Sylvain Joffre from the Finnish Meteorological Institute for his valuable comments and suggestions.

References

Bogren J, Gustavsson T, Karlsson M, Postgård U. 2000. The impact of screening on road surface temperature. *Meteorol. Appl.* **7**: 97–104.

Canadian Meteorological Centre. 2013. METRo 3.2.7 source code. http://home.gna.org/metro/ (accessed 29 January 2016).

Chapman L, Thornes JE, Bradley AV. 2001a. Modeling of road surface temperature from a geographical parameter database. Part 1: statistical. *Meteorol. Appl.* **8**: 409–419.

Chapman L, Thornes JE, Bradley AV. 2001b. Modeling of road surface temperature from a geographical parameter database. Part 2: numerical. *Meteorol. Appl.* **8**: 421–436.

Chapman L, Thornes JE, Bradley AV. 2001c. Rapid determination of canyon geometry parameters for use in surface radiation budgets. *Theor. Appl. Climatol.* **69**: 81–89.

Crevier LP, Delage Y. 2001. A new model for road-condition forecasting in Canada. *J. Appl. Meteorol.* **40**: 2026–2037.

Habrovský R, Tarjáni V. 2014. Kalman filter preprocessing within METRoSTAT project and application of the new method in the roadcast system. *Proceedings of the 17th International Road Weather Conference (SIRWEC)*, 30 January–1 February 2014, La Massana, Andorra. http://www.sirwec.org/Papers/andorra/33.pdf (accessed 29 January 2016).

Homleid M. 1995. Diurnal corrections of short-term surface temperature forecasts using the kalman filter. *Wea. Forecasting* **10**: 689–707.

Juga I, Nurmi P, Hippi M. 2013. Statistical modelling of wintertime road surface friction. *Meteorol. Appl.* **20**: 318–329.

Kangas M, Heikinheimo M, Hippi M. 2015. RoadSurf – a modelling system for predicting road weather and road surface conditions. *Meteorol. Appl.* **22**: 544–533.

Kršmanc R, Tarjáni V, Habrovský R, Slak AŠ. 2014. Upgraded METRo model within the METRoSTAT project. *Proceedings of the 17th International Road Weather Conference (SIRWEC)*, 30 January–1 February 2014, La Massana, Andorra. http://www.sirwec.org/Papers/andorra/26.pdf (accessed 29 January 2016).

Oke TR. 1987. *Boundary Layer Climates*, 2nd edn. Methuen & Co: Cambridge.

Sass BH. 1997. A numerical forecasting system for the prediction of slippery roads. *J. Appl. Meteorol.* **36**: 801–817.

Seity Y, Brousseau P, Malardel S, Hello G, Bénard P, Bouttier F, *et al*. 2011. The AROME-france convective-scale operational model. *Mon. Weather Rev.* **139**: 976–991.

Shao J. 1997. Improving nowcasts of road surface temperature by a backpropagation neural network. *Wea. Forecasting* **13**: 164–171.

Shao J, Lister PJ, Hart GD, Pearson HB. 1996. Thermal mapping – reliability and repeatability. In *Proceedings of the 6th International Road Weather Conference (SIRWEC)*, 17–19 April 1996, Birmingham, UK. http://www.sirwec.org/Papers/birmingham/34.PDF (accessed 11 March 2016).

Shao J, Thornes JE, Lister PJ. 1993. Description and verification of a road ice prediction model. *Transp. Res. Rec.* **1387**: 216–222.

Thornes JE. 1991. Thermal mapping and road-weather information systems for highway engineers. In *Highway Meteorology*, Perry AH, Symons LJ (eds). E & FN Spon: London; 39–67.

Vaisala. 2001. Vaisala Road/runway surface and depth sensor DRS511. Technical Note. http://www.vaisala.com/en/roads/products/roadweathersensors/Pages/DRS511.aspx (accessed 29 January 2016).

Vaisala. 2010. Vaisala remote surface temperature sensor DST111. Technical Note. http://www.vaisala.com/en/products/surfacesensors/Pages/DST111.aspx (accessed 29 January 2016).

Reprinted, with permission, from
Weather and forecasting, 32, no. 3, 991–1006,
doi: 10.1175/WAF-D-16-0158.1

II

Comparing the Performance of Two Road Weather Models in the Netherlands

VIRVE KARSISTO

Finnish Meteorological Institute, and Department of Physics, University of Helsinki, Helsinki, Finland

SANDER TIJM

Royal Netherlands Meteorological Institute, De Bilt, Netherlands

PERTTI NURMI

Finnish Meteorological Institute, Helsinki, Finland

(Manuscript received 29 August 2016, in final form 7 February 2017)

ABSTRACT

High-quality road condition forecasts are a prerequisite for road authorities to ensure wintertime road safety. Harsh winter conditions can cause problems for traffic not only in countries where snowy winters are common but also in regions where the temperature drops below the freezing point occasionally. This study reports on the evaluation of the Royal Netherlands Meteorological Institute's (KNMI) new road weather forecasting model by comparing it with the Finnish Meteorological Institute's (FMI) road weather model, both run for 321 Dutch road weather stations, four times daily (0300, 0900, 1500, and 2100 UTC) during the test period, 15 January–28 February 2015. Road surface temperature forecasts by both models were evaluated against observations. The KNMI model produced slightly more accurate forecasts than the FMI model. The main reason for the difference is probably due to the optimization of the physical properties of the KNMI model for the Netherlands, whereas the FMI model is designed for quite different Finnish wintertime meteorological conditions. However, in general the road surface temperature forecasts were of quite comparable quality.

1. Introduction

High quality road weather forecasts are needed to optimize wintertime road maintenance operations and services. The plowing and salting of roads consumes resources and are costly operations. As one example, around 100 million euros are spent annually for winter road maintenance in Finland (Venäläinen and Kangas 2003). A comparable amount is spent in the Netherlands even with much less frequent tough wintry weather conditions. Neglecting timely maintenance operations would lead to slippery roads, increasing the number of accidents, which would become even more expensive for society. In addition to injuries, casualties, and damaged vehicles, traffic congestion can cause long delays in transportation. Winter tires are not commonly used in the Netherlands, causing trucks to get stuck in steep access and exit areas of highways and blocking them under icy conditions. Salting and plowing can be planned well ahead and thus the costs can be minimized by making use of accurate road weather forecasts.

Many road weather models (hereafter RWMs) have been developed during the past 30 years (Rayer 1987; Jacobs and Raatz 1996; Chapman et al. 2001; Crevier and Delage 2001; Fujimoto et al. 2012; Yang et al. 2012; Kangas et al. 2015). The Finnish Meteorological Institute (FMI) initiated road surface temperature modeling activities in 1979 (Nysten 1980). The resulting model was in operational use during the early 1980s but was later discontinued. The model was also tested in the Netherlands within the European Cooperation in Science and Technology (COST) 30 bis project (David and Portal 1985). Data from road weather stations were collected, and an automatic system produced forecasts and warnings for a few hours in the future. The project also covered road/vehicle communications, automatic incident detection, and variable traffic signals.

The current operational RWM in FMI was developed in the late 1990s and has been operational since 2000

Corresponding author e-mail: Virve Karsisto, virve.karsisto@fmi.fi

DOI: 10.1175/WAF-D-16-0158.1

(Kangas et al. 2015). Several model improvements and developments have been made thereafter, including a pavement condition forecast application for pedestrians (Kangas et al. 2015) and a perfect prog-type statistical application to forecast road surface friction (Juga et al. 2013).

The Royal Netherlands Meteorological Institute (KNMI) RWM model was developed during 2014–15. This paper reports on the assessment of forecast quality of this brand new model by comparing its output with the FMI RWM, which has a long history of operational use. Both models' results were evaluated against road surface temperature observations. Model comparisons can be truly beneficial in finding out good properties as well as weaknesses of the models, which then leads to potential model improvements. There have been very few earlier comparative studies like this work despite the relatively high number of RWMs in use. Thornes and Shao (1991) compared three ice prediction models developed in the United Kingdom: the ICEBREAK model (Shao and Lister 1996), the Met Office model (Rayer 1987) and Thornes' model (Thornes 1984). All of these three models used the same input data and were run for a single test site in 24-h cycles. The ICEBREAK model showed the best performance based on model bias, standard deviation, and root-mean-square error.

Having a separate RWM in addition to a general numerical weather prediction (NWP) model is important since the physical processes can be evaluated in more detail with separate models. An RWM can model conditions specific to the road surface, whereas NWP uses the generalized land-use types. In addition, the effect of traffic on the amount of water and snow can be taken into account in an RWM. Since RWMs are usually one-dimensional, they can also use observations made at certain road points in their initialization rather than using interpolated observations. The present study compares the road surface temperature forecasts made by the KNMI and the FMI road weather models. Both models used the same observations and NWP forecast data as input to make the results comparable. The aim of the study is to assess the performance of the new KNMI model using the FMI model as a reference. Therefore, other road weather models are not included in the present study, but comparison with other models could be an important research topic in the future. Section 2 defines the physical and technical properties of both the FMI and KNMI models. Section 3 introduces the observations and the forcing datasets. The results are reported and analyzed in section 4 using standard verification metrics, and section 5 concludes the paper with the final discussion. Finally, the appendix gives a more detailed description of the physical equations used in the models.

2. Model descriptions

a. Initialization

Both the KNMI and FMI RWMs are one-dimensional heat balance models that require as their input forecasted parameters from a three-dimensional NWP system. The input includes the following parameters interpolated to the respective road points: air temperature, dewpoint temperature/relative humidity, wind speed, incoming long- and shortwave radiation, and precipitation. The models also make use of observations from road weather stations (RWSs) when defining the models' initial temperature profiles. However, the two models adapt different procedures in determining the initial state of the forecast. The FMI model is run for 2 days prior to the latest measurements using observations from road weather stations as the forcing. The temperature of the first surface layer is set to be the observed road surface temperature at each time step, and the temperature profile evolves according to a heat transfer equation. The model then includes a 3-h period during which the forecasted radiation is adjusted to the observed road surface temperature. This method is called *coupling* and is explained in more detail by Crevier and Delage (2001) and Karsisto et al. (2016). The coupling phase starts at the time when the input forecast from the NWP model is initiated. The model calculates the temperature profile during this phase based on the heat balance equation using the NWP forecast as the forcing. The method determines a correction coefficient iteratively either for forecasted longwave (LW) or shortwave (SW) radiation so that at the end of the period the forecasted road surface temperature fits the observed road surface temperature. This correction coefficient is consequently used during the actual forecast phase. It approaches exponentially unity (1.0) as the forecast evolves and, after 6 h, the correction is typically about 20% of its original value. The correction coefficient is given for the radiation variable that has a higher intensity at the end of the 3-h period.

The KNMI model applies a quite different initialization procedure. The model run starts 1 h before the beginning of the actual RWM forecast, and the initial temperature profile is taken from the previous forecast rather than running the model with observations. However, the profile is adjusted according to the temperature difference between the observed surface temperature and the modeled surface temperature from the

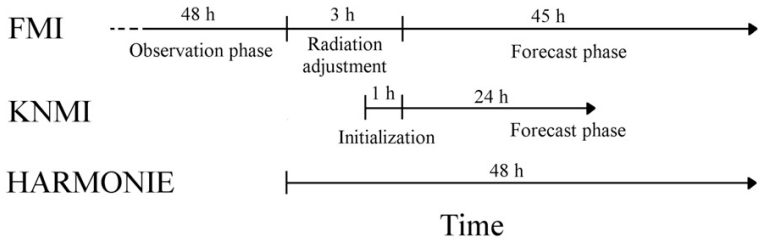

FIG. 1. The phases of FMI and KNMI model runs and their relation to the driving HARMONIE forecast.

previous forecast. The adjustment is 100% at the top layer and decreases linearly to 0% at the bottom layer. Then, the model also has a period that is used to correct the forecasted radiation according to the forecasted road surface temperature, but the length of the period is 1 h and the adjustment is not done iteratively. Instead, it is based on general calculations of how much energy is needed to change the surface temperature. The model is run for 1 h using the latest available forecast from an NWP model as the forcing, and the forecasted surface temperature is compared with the observed temperature at the end of the period. If the difference is more than 0.05 K, the model calculates coefficients either for LW or SW radiation based on the general calculations, so that the change compensates for the temperature difference. The coefficient is given for SW radiation if its intensity is larger than 100 W m^{-2} in the 1-h period and otherwise for longwave radiation. In the actual forecast phase the chosen radiation parameter is corrected using this coefficient. However, the correction used is only 50% of the original coefficient, because the adjustment of the 100% correction did not yield results that were as good as in the sensitivity tests. The correction coefficient remains the same for 3 h and after that is scaled linearly back to 1.0 in 9 h. The temperature profile is modified again before the start of the forecast phase using an observed surface temperature similar to that at the start of the model runs.

Due to the different initialization methods, the model runs were organized in such a way that the actual forecast phase started at the same time with both the KNMI and FMI models (Fig. 1). The necessary input forecasts were taken from the HIRLAM–ALADIN Research on Mesoscale Operational NWP in Euromed (HARMONIE; Bengtsson et al. 2017) model run by KNMI. Further details about this version of HARMONIE are given in section 3a. The FMI model requires 3 h during which observations and forecasts are available simultaneously. The starting time of the actual forecasts is therefore always 3 h after the HARMONIE starting time. To get the same starting time for the KNMI model, the model run must start 2 h after the HARMONIE forecast run, because the initialization period of the KNMI model is only 1 h. For example, if a HARMONIE forecast starts at 0000 UTC, the actual forecast phase in both models begins at 0300 UTC. The forecast length was 45 h for the FMI model and 24 h for the KNMI model.

b. Physical properties

The physical properties of the surface and the road are quite different for the KNMI and FMI models (Table 1). The ground is divided into separate layers in both models, and the heat transfer is calculated between each layer at each time step. The KNMI model has 20 layers and the FMI model 16. The first two layers in the FMI model are considered to be asphalt and the rest have soil properties, whereas all of the layers are considered to be asphalt in the KNMI model. The layers of the KNMI model are much thinner than the FMI model layers, and the thickness of the road surface layer is only 0.3 cm, when it is 1.5 cm in the FMI model. The difference is highlighted in Fig. 2. The output surface temperature in the KNMI model is given as the temperature of the uppermost layer, whereas the FMI model uses the average of the top two layers as the output temperature. The depth of the lowest layer in the KNMI model is 0.33 m, but it is much deeper in the FMI model, with the middle point of the bottom layer being as deep as 4.28 m in the ground. The FMI model has a relatively long initialization period partly because it takes time to adjust the temperature of the lower layers. The density and specific heat of asphalt are larger in the KNMI model than in the FMI model, compensating for the thinner model layers. Also the asphalt heat conductivity is higher in the KNMI model. Moreover, the KNMI model has a separate mode for bridges, in which the lowest model-layer temperature is influenced by the air temperature. This mode was used when running the model for road weather stations on bridges.

The density, specific heat, albedo, absorption, and emissivity parameters of the KNMI model were determined before the start of the comparison study by

TABLE 1. The surface and ground properties of the FMI and KNMI models. The KNMI model has only an asphalt layer and thus the soil properties are not given in the table. In addition, the KNMI model does not use porosity in the model calculations.

	FMI (Kangas et al. 2015)	KNMI
No. of ground layers	16	20
Depth of the bottom layer (m)	4.28	0.33
Thickness of the first layer (cm)	1.5	0.3
Specific heat, asphalt ($J\,kg^{-1}\,K^{-1}$)	919	2000
Heat conductivity, asphalt ($W\,K^{-1}\,m^{-1}$)	0.5	1.25
Density, asphalt ($kg\,m^{-3}$)	2110	3000
Porosity, asphalt	0.1	—
Specific heat, soil ($J\,kg^{-1}\,K^{-1}$)	813	—
Heat conductivity, soil ($W\,K^{-1}\,m^{-1}$)	1.4	—
Density, soil ($k\,gm^{-3}$)	1600	—
Porosity, soil	0.4	—
Surface albedo	0.1, bare road; 0.6, snow; 0.1–0.6, ice	0.20
Emissivity	0.95	0.94
Absorption of long wave radiation	0.95	0.99
Roughness length for momentum (m)	0.4	0.001
Roughness length for heat (m)	0.001	0.001
Temperature reference height (m)	2	2
Wind reference height (m)	10	10

performing sensitivity tests for roughly 15 stations during the time period 1–28 February. Different parameter values were tested to find the best combination. Varying the parameters has a significant effect on the model bias values; for example, the negative road surface temperature bias for runs starting at 1800 UTC increases from −0.8° to −1.3°C at 0600 UTC (+12-h forecast) when the density is decreased from 3000 to 2000 $kg\,m^{-3}$. The 1200 UTC (+18-h forecast) positive bias increases from 1.3° to 2.0°C. These tests were done without initial surface temperature correction. As a result of the optimization of the physical properties, the KNMI parameters are heavily tuned toward values that correspond to observations. The high density and heat capacity values make the KNMI model slower to react to radiation and air temperature changes than it would be otherwise. This aims to correct for the effects of shading, which greatly affects the road surface temperature (Bogren et al. 2000). The corresponding parameters are defined in the FMI model in an attempt to produce reliable results for the Finnish roads. The KNMI model uses a sky-view factor of 0.9 that is the same at all locations, because not enough sky-view factor data were available at the time of this project. The option to use a station-specific sky-view factor is available in the KNMI model but not been used yet. The FMI model did not use the sky-view factor in this study. Some sensitivity tests were performed with the KNMI model to estimate the effect of the sky-view factor on the surface temperature. The model was run for the second half of February 2016 with sky-view factors varying from 0.4 to 1.0. Decreasing the sky-view factor to 0.1 caused the model bias to be 0.2–0.3 K more positive in the forecasts that started at 1800 UTC after 10 h of nighttime running. It must be noted that the test was run at lower density and conductivity values for asphalt than in the operational model, which caused the model to be a bit more sensitive.

c. Output

The main output variable in both models is the road surface temperature. In the KNMI model it is the temperature at 1.5-mm depth inside the road, whereas it is the average temperature of the top two layers with thicknesses of 1.5 and 3.25 cm in the FMI model. The ground temperatures at depths of 3 and 20 cm are also produced by the KNMI model. The whole temperature profile of the FMI model was saved every hour for further analysis during this study. However, only the temperatures at depths of 3.0 and 6.5 cm are produced by the FMI model under normal operational forecast

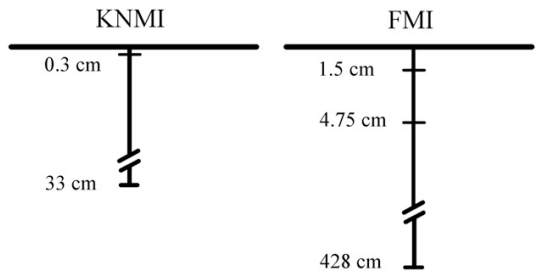

FIG. 2. The lower boundaries of the surface layers and the middle-point heights of the bottom layers in the KNMI and FMI models.

conditions. Both models also provide estimates of the amounts of water and ice on the road. The FMI model produces, in addition, separate values for snow and frost, whereas they are all considered as ice output in the KNMI model. The snow and ice also have some effect on the surface temperature values, causing them to remain near zero in the models during melting. There were some cases during the test period when snow was forecasted in the models and also at least one case when there was actually snow on the roads.

A road surface condition index is an additional output value of both models defining whether the road is dry, wet, icy, snowy, etc. Moreover, the sensible and latent heat fluxes and the net surface radiation were included as model output. The FMI model does not normally produce these variables but they were included in this study to enable more detailed comparison between models. All output parameters were produced every full hour. The FMI operational model version also calculates the surface friction (Juga et al. 2013) as well as indices for pedestrians and drivers depicting whether the conditions are normal, difficult, or very difficult (Kangas et al. 2015).

3. Data

a. NWP forecast

The high-resolution HARMONIE model is based on the AROME model developed by Météo-France and is described in more detail by Seity et al. (2011) and Bengtsson et al. (2017). HARMONIE has been in operational use at KNMI since summer 2012 (Baas and Van den Brink 2014), where the model domain extends roughly from 42° to 60°N and from 10°W to 17°E for an area of $2000 \times 2000\,\text{km}^2$. HARMONIE version 36h1.4 with a resolution of 2.5 km was used in this study during 15 January–28 February 2015 to provide input forecasts for both RWMs. Within HARMONIE, three-dimensional variational data assimilation (3DVAR), in addition to blending of the large-scale High Resolution Limited Area Model (HIRLAM; Undén et al. 2002), is used to improve the initial conditions in the atmosphere. More information about data assimilation can be found in the work by Seity et al. (2011) and Bengtsson et al. (2017). HARMONIE uses a separate externalized surface model (SURFEX) library (Masson et al. 2013) to model surface processes. The SURFEX library uses four tile types (land, town, sea, and inland water) to describe the grid-box area and the physical parameterizations used are different for each type. Output values are calculated as weighted averages of the results for each tile according to their relative areas in the grid box. The parameterizations used are described in more detail by Masson et al. (2013), Seity et al. (2011), and Bengtsson et al. (2017).

The HARMONIE runs were initiated at 0000, 0600, 1200, and 1800 UTC each day. Because the initialization procedure of the FMI model takes 3 h, the RWM forecasts started at 0300, 0900, 1500, and 2100 UTC. The local time in the Netherlands is UTC + 1 h in winter, meaning that the starting hours correspond to 0400, 1000, 1600, and 2200 local time (LT), respectively. In total there were over 50 000 forecasts considered when all road weather stations were included; that is a large enough dataset to determine the behavior and the quality of the different models. The studied period contains multiple days with a large daily temperature cycle as the sun rises high enough in the sky to cause significant heating of the surface. Many days in February were very sunny, but there were also several cases during the test period with very cloudy conditions. The minimum temperatures in De Bilt, near the center of the Netherlands, were around −5°C and the maximum temperatures were around 10°C during the test period. In the center of the Netherlands there were more than 20 days with a minimum 2-m temperature below 0°C. In De Bilt there was one case of freezing rain turning into snow on 24 January and one case of light freezing rain on 7 February. This gives enough of a variety of conditions for the models to be tested, allowing for reactions to the daytime heating and nighttime cooling and the behavior for temperatures around 0°C.

b. Observations

Observation data were obtained from 321 road weather stations scattered across the country and maintained by the Dutch Rijkswaterstaat. Each station provides up to 12 road surface temperature sensors and 12 conductivity sensors at a single location. The sensors are typically located at slightly different places near the station (e.g., on different lanes). The surface temperature sensors are installed 2 mm below the surface, which is close to the middle point of the uppermost KNMI model layer at 1.5 mm. The stations also measure air temperature, dewpoint temperature, and, at some locations, soil temperature. The observation frequency is 5 min. Before producing the forecasts, the observations underwent an automatic quality control procedure to remove suspicious values. In total there were 298 stations where the RWMs could be run with proper initialization.

Because of several surface temperature observations being available at one location, there was the need to decide which sensor should be used as the RWM input. Road surface temperatures can vary significantly across the width of a road profile (Chapman and Thornes 2011). It is most relevant for the road maintenance authorities to get information on the overall lowest surface temperature at all locations to be able to determine

potential areas prone to freezing. The model can be adjusted to best predict the minimum temperature value at the station area when initialized with the lowest temperatures. Consequently, data from the sensor with the lowest temperature were used in the model initialization procedure, and this temperature was selected at each observation time. Therefore, the input data can include observations from several different sensors rather than being a full time series originating from one single sensor. The differences between sensors are usually less than 2 K during nighttime, but can be as large as 6 K at noon, highlighting the effect of the station's location on road surface temperatures.

4. Results

a. Comparison between HARMONIE and KNMI RWM

Before focusing only on the RWMs, the error statistics of the surface temperature forecasts made by the KNMI RWM and HARMONIE model were compared. Around 15 road weather stations were selected and the 0000 UTC forecast runs were analyzed for the period 1–28 February. The HARMONIE model has a negative surface temperature bias from around −0.5 to −1.0 K throughout the forecast, whereas the KNMI model bias is mostly positive except during the morning, when the most negative value is about −0.3 K. The KNMI model's positive bias varies from around 0.1–0.3 K during the nighttime to 0.4–1.0 K during the daytime. In the RWM the heat fluxes and ground properties are specified for the road, which explains the smaller bias values during the night. However, the asphalt seems to become too warm during the day in the model.

During the first forecast hours the root-mean-square error (RMSE) values of the KNMI RWM are more than 1 K better than those of the HARMONIE model. This considerable difference is expected because the RWM uses road surface temperature observations in the initialization. During the daytime the RMSE difference between models is around 0.0–0.3 K, but the difference increases again during the evening, and the KNMI RWM has considerably better RMSE values throughout the rest of the 24-h forecast. Overall, the HARMONIE model can predict the afternoon temperatures a little better than the KNMI RWM, but the KNMI model is considerably better during the rest of the day.

b. Bias

This and the following sections contain the verification results of the road surface temperature forecasts from the model runs starting at 0300 and 1500 UTC,

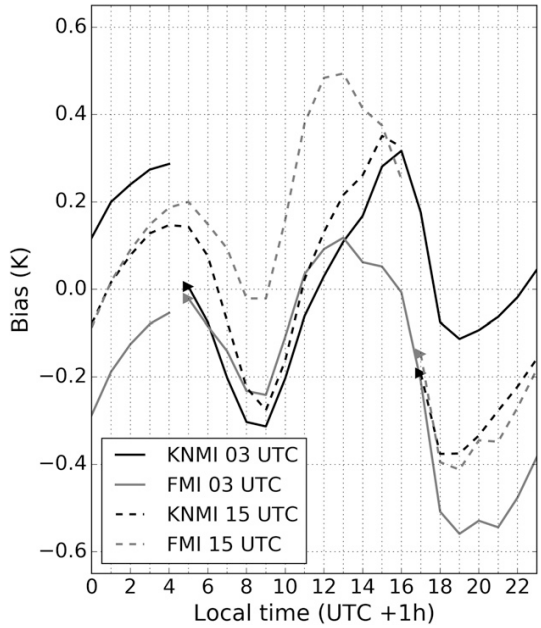

FIG. 3. Bias of road surface temperature as a function of local time for 0300 UTC (continuous line) and 1500 UTC (dashed line) forecast runs. Values are calculated for all stations and for the whole test period (15 Jan–28 Feb 2015). KNMI (FMI) results are shown by the black (gray) line. The triangles show the first forecast hour.

because they are the most relevant values for road maintenance (like salting the roads). Bias values were calculated separately for each forecast hour (Fig. 3). The forecasted values are very close to each other during the first 8–12 h from the start of the forecasts, with differences of less than 0.1 K. The differences become larger beyond 9 h when the biases show different signs.

The bootstrap method was used to determine whether the bias difference between the models is statistically significant (Efron and Tibshirani 1993; Hogan and Mason 2011). Some 10 000 bootstrap samples were generated with replacement for each lead time using the sample size of the original data. Bias values of both models' were calculated from each sample. Then, the 1st and 99th percentiles were determined from the distribution of the bias differences between the models. If zero is not included in the obtained range, the differences between the models are considered to be statistically significant (corresponding to a p value of 0.02). For the forecasts starting at 0300 UTC, the differences are statistically significant for all times shown in the Fig. 3, except for the 1300 LT forecast. However, for the 0600 LT forecast the significance level is just barely attained. For the 1500 UTC forecasts the results are statistically

significant for all times except for 2000, 0000, 0100, and 0200 LT.

The model biases have a daily cycle with the largest positive values during the day around 1300–1600 LT and during the night around 0500 LT, reaching their lowest negative values in the morning at 0900 LT and in the evening at 1900 LT. One reason for the high positive daytime biases is the shading effects, which are not taken into account in the models. This causes significant temperature overestimation at a number of stations during the time of the dominant shortwave radiation. Ignoring the sky-view factor may be one fundamental reason for the negative nighttime bias, because the longwave radiation from the surrounding objects is not taken into account in the model. The midday temperature bias maximum shows up about 3 h later in the KNMI results than in the FMI model results. The bias maximum tends to occur at the same time as the surface temperature maximum in the FMI model, but in the KNMI model it occurs after the temperature has started decreasing. This means that the KNMI model usually cools more slowly during afternoon than the FMI model. This must be considered to be a net effect of the differences in the model physical properties, and it is hard to find an individual reason causing the behavior. The considerably thinner surface layers in the KNMI model would presumably lead to faster cooling than in the FMI model, but the results show that other differences between the models are more dominant. First, the slower cooling in the KNMI model is supported by the larger heat capacity of the road material. Second, the sensible heat fluxes and latent heat fluxes tend to be more negative in the FMI model during the daytime as a result of the larger roughness length for momentum, and also the net radiation has smaller values, which supports faster cooling. Consequently, the slower reaction in the KNMI model must be considered to be net effect of the heat capacity, conductivity, layer thickness, and differences in the model fluxes. More information about the fluxes and their calculation in the models can be found in the appendix.

As mentioned above, the daytime maximum temperatures tend to be too high in both models. In addition, the nighttime minimum temperatures tend to be too warm, except the FMI 0300 UTC forecasts, where the bias remains on the negative side. One major difference between the models is that the FMI model becomes much more negative in the early evening in the 0300 UTC run, and it remains negative throughout the night. Again there are multiple factors causing this difference between the models. One reason could be that in the KNMI model the heat stored to the ground during the daytime is transferred more efficiently to the surface during nighttime as a result of the larger heat conductivity value of the asphalt. Also, the net radiation and latent heat flux are less negative in the KNMI model during the nighttime in general.

Contrarily, the KNMI model 1500 UTC run has a more negative road surface temperature bias at 0900 LT, whereas the FMI model bias is close to zero. The KNMI model seems to react more slowly to the increasing shortwave radiation during the early morning than does the FMI model. This must be considered again to be a net effect of the model differences. The behavior is partially caused by the larger heat capacity of the surface material in the KNMI model, although the results contradict the fact that the surface layers are thinner in the KNMI model. In addition, the sensible heat flux in the FMI model is larger during the early morning, which supports the faster warming. However, it needs to be highlighted that the values are averaged over all stations, and that there are huge variations between stations even within the relatively small area of the Netherlands. As an example, Fig. 4 shows the biases of one forecast hour calculated separately for all stations. The values reveal that although the FMI model bias is about zero in Fig. 3, there are in reality many stations with either positive or negative biases canceling each other. The KNMI model biases are mainly negative, but there are also stations with positive values.

A part of the bias in the RWMs is caused by the errors in the HARMONIE forecasts. Their effects could not be validated thoroughly because of the lack of the radiation and cloudiness observations at the road weather stations, but the effect of removing the 2-m temperature bias was tested by running the KNMI model with observed values. This reduced the nighttime bias by 50%, so the HARMONIE forecast errors clearly have a significant effect on the accuracy of the RWM forecasts.

c. *RMSE*

Figure 5 shows the RMSE for the same road surface temperature forecasts as in Fig. 3. The FMI model has on average a slightly larger RMSE than the KNMI model but the differences are mostly around 0.1 K. The differences were determined to be statistically significant with all lead times using the same bootstrap method described in the previous section. The RMSE values are greatest at midday, and the difference between the models grows to approximately 0.3 K in the 0300 UTC run and up to 0.5 K in the 1500 UTC model run. Daily maximum temperatures are usually difficult to predict, because they depend so much on the total radiation budget. In addition, the observational data used in the verification originates from a sensor giving the lowest temperature at each station. It is the sensor that,

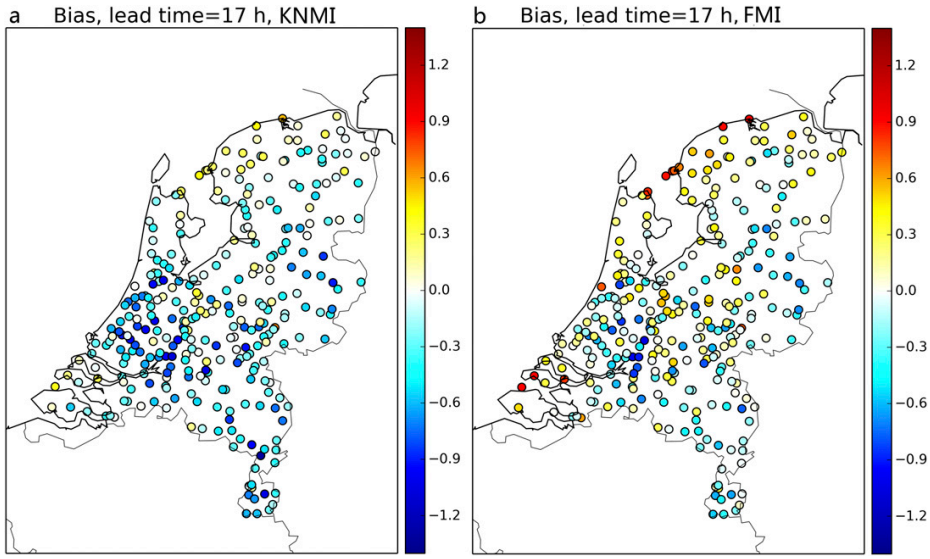

FIG. 4. Biases (K) of road surface temperature forecasts in the Netherlands for the (a) KNMI and (b) FMI models at 0900 LT. The forecasts were initiated at 1500 UTC and the lead time is 17 h, averaged over the whole time period from 15 Jan to 28 Feb 2015.

on average, has the largest influence of shading, which is not taken into account in the RWMs. This produces larger RMSE values at midday than would be obtained from observational data consisting of the maximum observations among the sensors. There are small reductions in the RMSEs at around 1000 and 1800 LT. The 0300 UTC model run produces lower RMSE values than the 1500 UTC model run for forecast lengths of a few hours. The reason can be the radiation adjustment. The radiation changes rapidly around 1500 UTC (1600 LT), so the radiation correction factor determined during the initialization does not fit that well during later hours in either model. During the early morning, at 0300 UTC (0400 LT) the radiation does not change that much and the correction is more appropriate. When studying other model runs, it was noted that forecasts initiated at 0900 UTC produced the largest RMSE values in the short range, the error being approximately 0.8 K in the first forecast hour. This is reasonable because it is hard to give an accurate radiation adjustment because of the unsteady radiation around 0900 UTC (1000 LT). Similarly to the bias values, there was much variation in the RMSE values between individual stations.

d. Categorical performance

One of the most important issues in road weather forecasting involves making accurate predictions around 0°C. To verify this, the hit and false alarm ratios were computed within five different temperature ranges as follows: $T < 0$°C, $-5.0° < T < -1.0$°C, $-1.0° < T < 0.0$°C, $0.0° < T < 1.0°$, and $1.0° < T < 5.0$°C, where T refers to the road surface temperature. The whole dataset was categorized utilizing the common contingency table shown in Table 2, followed by the computation of the probability of detection (POD) and the false alarm ratio (FAR) within these categories (WMO 2014):

$$\mathrm{POD} = \frac{a}{a+c} \quad \text{and} \tag{1}$$

$$\mathrm{FAR} = \frac{b}{a+b}. \tag{2}$$

The POD defines how frequently an event is correctly forecasted in relation to the number of cases when the event is observed. The FAR, on the other hand, indicates the number of false alarms in relation to the number of cases when the event is forecasted. The results are shown collectively in Fig. 6 in the form of a categorical performance diagram (Roebber 2009; Ebert et al. 2013). The total number of forecast cases was 13 000. The y axis shows the POD values and the x axis the FAR values in a reversed scale. A perfect forecast would fall in the top-right corner of the diagram, where POD = 1 and FAR = 0. The dotted lines represent the frequency bias $[(a + b)/(a + c)]$, which describes whether there was over- or underforecasting of the event in the given category. Values higher (lower) than 1 indicate overforecasting (underforecasting). Figure 6

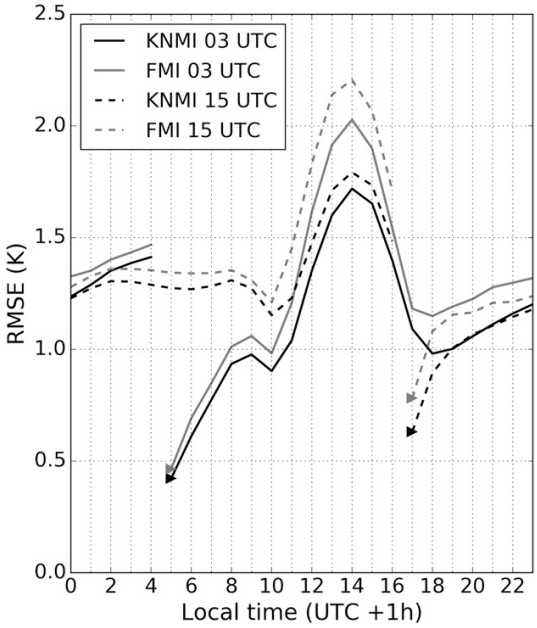

FIG. 5. RMSE (K) of road surface temperature forecasts as a function of local time for 0300 UTC (continuous line) and 1500 UTC (dashed line) forecast runs. KNMI (FMI) results are shown by the black (gray) line. The triangles show the first forecast hour.

further shows, with the continuous line, the so-called critical success index [CSI; $a/(a + b + c)$], which expresses the relation of hits to the total number of cases where the event was either observed or forecasted. In an ideal case, CSI would be equal to 1. The error bars in Fig. 6 represent the 95% confidence interval that is calculated using the error variance as described by Hogan and Mason (2011).

Figure 6 highlights that the scores for both of the models are quite similar. The same bootstrap method as described in section 4b was used to find out the statistical significance of the differences. The FMI model has typically slightly larger FAR values than the KNMI model. The differences were significant except for the forecasts started at 0300 UTC in the range $-1.0° < T < 0.0°$C. The differences in POD values were statistically significant only for temperature ranges below 0°C of the 0300 UTC model run and for the range $1.0° < T < 5.0°$C of the 1500 UTC model run. In the 0300 UTC model runs the KNMI model has a somewhat higher POD for ranges $T < 0.0°$C and $-5.0° < T < -1.0°$C, but the POD of FMI is slightly better in the range $-1.0° < T < 0.0°$C. In the 1500 UTC model run the KNMI model has a little higher POD than the FMI model with range $1.0° < T < 5.0°$C. The scores for a larger hit range give better results than scores calculated for a range of 1°C, because the probability of a correct forecast is higher with a larger temperature range. Within ranges $-1.0° < T < 0.0°$ and $0.0° < T < 1.0°$C, both the POD and FAR results are around 0.5 for the 0300 UTC run, and the verification markers are even closer in the bottom-left corner for the 1500 UTC run, indicating lower forecast quality. Moreover, the 0300 UTC forecasts produce in general better results than the forecasts initiated at 1500 UTC. This is excepted because the surface temperature usually varies more between 1500 and 1800 UTC in the Netherlands than between 0300 and 0600 UTC, and thus the values from the 0300 UTC model run do not differ that much from the observations used in the initialization and are easier to predict.

TABLE 2. The contingency table.

		Event observed	
		Yes	No
Event forecast	Yes	*a*	*b*
	No	*c*	*d*

Some of the POD and FAR values are quite dependent on the time of day, as was the case with the bias and RMSE. This can be seen in Fig. 7, which represents the mean POD and FAR values as a function of local time. In the temperature range $1.0° < T < 5.0°$C, the smallest POD value and the largest FAR value are detected at 1400 LT. This is in agreement with the RMSE values, where the maxima were also found around midday as a result of the difficulties in predicting the daily maximum temperatures. This feature cannot be seen within temperature ranges $T < 0.0°$ and $-5.0° < T < -1.0°$C, since there were so few observed and forecasted values at midday that the POD and FAR values could not be calculated. Instead, these temperature ranges are overpredicted in the evening, when there is a peak in the FAR values. This may also be seen in Fig. 6, where the 1500 UTC runs with both models produced relatively large FAR values within these categories. Both models have a cold bias in the evening, so the reason for this behavior is probably that the surfaces in the models cool too fast. During the nighttime the FAR is considerably smaller. The time dependency is not clear within temperature ranges $-1.0° < T < 0.0°$C and $0.0° < T < 1.0°$C, in which both the FAR and POD values are worse compared to all other temperature categories.

e. *Relative difference between models*

The median differences of the surface temperature forecasts of the two models were finally analyzed to better understand the dissimilarities in their performance. Figure 8 shows the results. Overall, the differences are

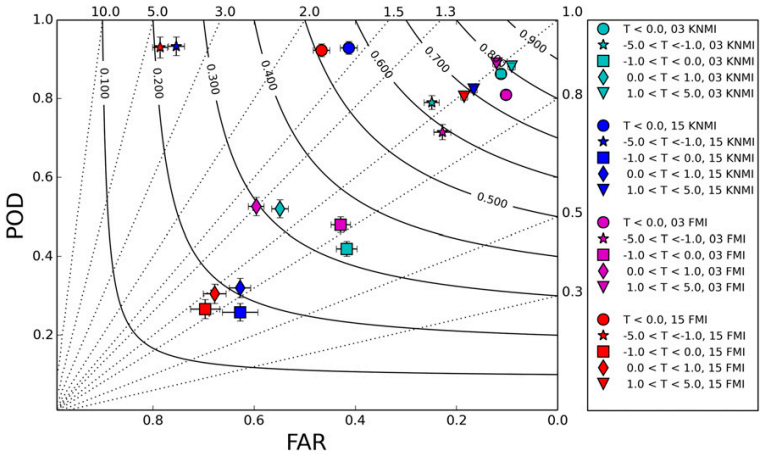

FIG. 6. Categorical performance diagram for forecasts initiated at 0300 and 1500 UTC for 3-h lead times. KNMI results are shown with cyan markers for 0300 UTC forecasts and with blue markers for 1500 UTC forecasts. FMI results for the same initialization times are shown with magenta and red markers, respectively. Circles represent results for $T < 0$°C, stars for the range $-5.0° < T < -1.0$°C, squares for $-1.0° < T < 0.0$°C, diamonds for $0.0° < T < 1.0$°C, and triangles for $1.0° < T < 5.0$°C. FAR runs along the x axis and POD along the y axis. Dotted lines represent the frequency bias and continuous lines the CSI.

relatively small, and the median absolute difference is always less than 0.7 K. The median difference is close to zero during the first 8 h of the 0300, 1500, and 2100 UTC forecast runs. However, in the 0900 UTC run the KNMI model is relatively warmer than the FMI model at the beginning of the forecast. The radiation changes rapidly in the morning and, consequently, the different initialization methods generate larger differences between the models. Results also show that the FMI model is usually a bit warmer in the morning for the 0300, 1500, and 2100 UTC runs, and the difference is largest at 0900 UTC. The surface temperature in the KNMI model usually rises more slowly in the morning, which is seen also in the model bias results and is caused as a net effect of the many model differences, as discussed in the section 4b. As the day advances, the KNMI model becomes warmer, and the difference becomes largest in the evening around 2000 UTC. It was seen also in the bias results that the FMI model tends to be colder during the nighttime, and the reasons for this were also discussed in the section 4b.

The standard deviations of model differences were also calculated (Fig. 9). The results follow the same pattern as for the RMSE, being largest around 1300 UTC and dropping around 0800 and 1700 UTC. A comparison of Figs. 8 and 9 shows that the KNMI model is usually a little warmer than the FMI model at midday, but the discrepancy between forecasts is large. In other words, there are also many cases where the FMI model is warmer during daytime. In the morning, when the KNMI model is generally colder, the standard deviation is smaller, so there are relatively fewer cases when the KNMI model is warmer in the morning. Correspondingly, it is not very common for the FMI model to be warmer than the KNMI model in the evening.

5. Discussion

The quality of the new (2015) KNMI road weather model was assessed by comparing it with the well-established road weather model of the FMI. The KNMI model generated somewhat smaller forecast errors across the Netherlands than the FMI model, confirming the applicability of its operational use for Dutch highways. The reason for the somewhat better performance is its optimization of the physical properties of local Dutch roads. The FMI model, on the other hand, has been designed by default for Finnish roads, whose physical properties are not considered totally suitable for the Netherlands. This study highlights the importance of the optimization of model physical properties when being implemented in new climatological and physiological environments. In the Netherlands the asphalt properties may vary across different areas; so, further studies are needed where physical properties are individually optimized for relevant road weather stations. Overall, the surface temperature forecasts of the models are quite similar, although the surface-layer thicknesses are very different in the two models. The

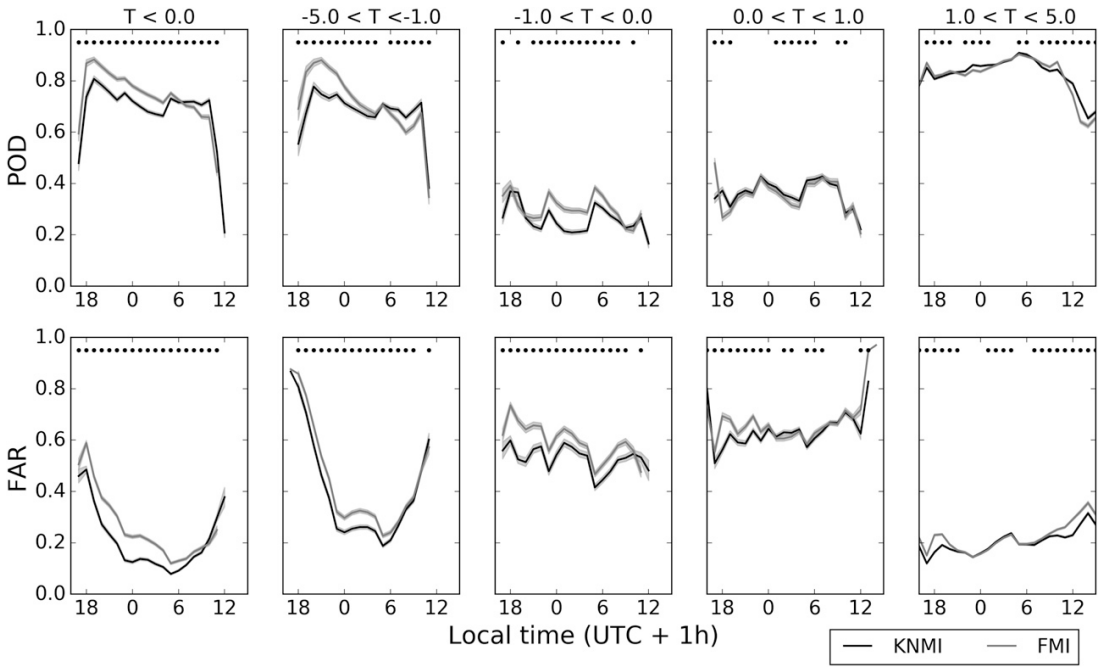

FIG. 7. Mean (top) POD and (bottom) FAR as a function of local time in the five different temperature range categories. The time values are from 1600 LT to 1500 LT from left to right. Averages are taken from the POD and FAR values calculated for 0300, 0900, 1500, and 2100 UTC model runs, i.e., each value corresponds to the same time of day regardless of the lead time. The means for the KNMI (FMI) model are shown by black (gray) lines. Not shown are cases where there were fewer than 100 hit values in the *a* category for POD and, correspondingly, fewer than 100 miss values in the *b* category for FAR. The 95% confidence intervals are shown by gray shading. The black dots show the cases where the differences between the models were statistically significant at the 98% confidence level.

net effect caused by the differences in the heat fluxes and physical parameters like asphalt heat capacity caused the KNMI model to react more slowly to the temperature changes during the morning and evening, despite the fact that the thickness of its surface layers was much thinner than those of the FMI model.

The use of the lowest surface temperature measurements at each station made the forecasting of daily maximum temperatures a challenge, since the possible shadowing effects at these locations can make the surface colder than forecasted. Shadowing was not taken into account except in the initialization process, because shadow factors have not been determined for the station locations. Doing this would have been too time consuming of a task in the present context. In the KNMI model the optimal values for each station are currently tested by running simulations with different heat conductivity and sky-view factor values. It is planned that in the future the sky-view and shading factors would be determined from a very high-resolution (25 cm) height map of the Netherlands. In the FMI the use of sky-view factors is currently tested in the in a small area of Norway as part of the Advanced Snow Plough and Salt Spreader Based on Innovative Space Technologies—Winter Road Maintenance (ASSIST WRM) project, where they are determined from 100-m-resolution height maps. Plans include testing different heat capacity and conductivity values for Finnish road weather stations to find the best combinations.

To develop RWMs, comparing results from different models is highly beneficial. However, there are very few recent published road weather model comparison studies. Thornes and Shao (1991) stated in the beginning of the 1990s that commercial reasons prevented the comparison of other than the three models that were analyzed in their study. However, thanks to the development of communication networks and scientific collaboration, it is now easy to share large datasets between countries. It has become possible for collaborating institutes to run their models with commonly shared input data and without necessarily providing access to local model codes if that should be the case for preventing collaboration. Further comparison studies similar to what has been performed here but with more participants would be highly interesting. However, even with only two partners both parties benefited greatly

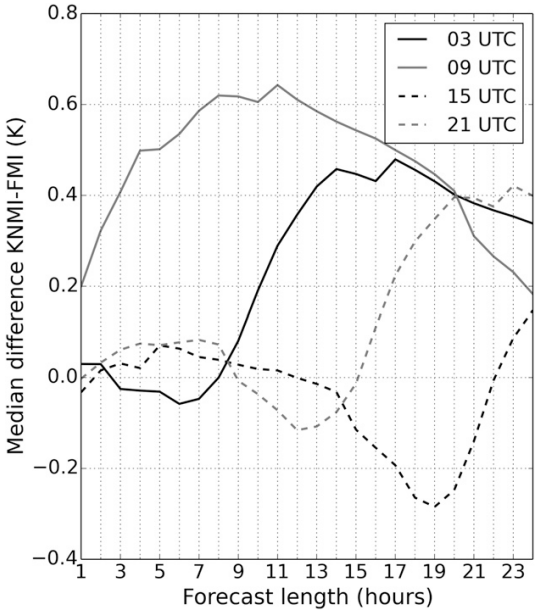

FIG. 8. Median difference of road surface temperature forecasts between KNMI and FMI models (KNMI − FMI) as a function of forecast length. Included are all stations where road surface temperature observations were available during the full test period (15 Jan–28 Feb 2015). Forecasts initiated at 0300 UTC are shown by the continuous black line, the 0900 UTC run by the continuous gray line, the 1500 UTC run by the dashed black line, and the 2100 UTC run by the dashed gray line.

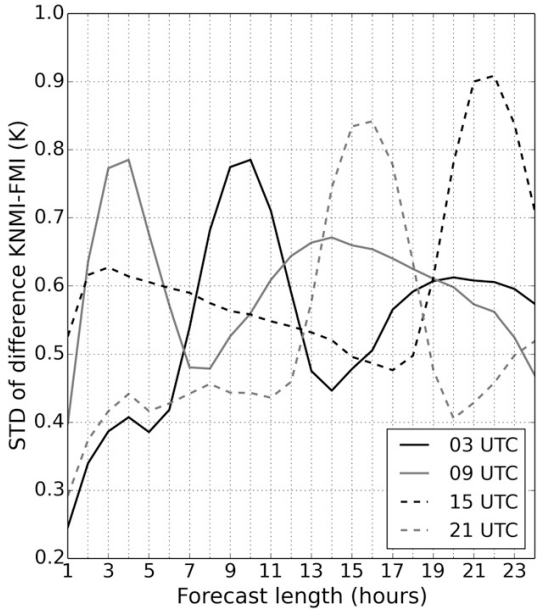

FIG. 9. As in Fig. 8, but for the standard deviation (STD).

from the collaboration, gaining valuable guidance and information for further development of their local weather models.

Acknowledgments. We thank Suomen Kulttuurirahasto (the Finnish Cultural Foundation) for financial support of road weather forecasting research.

APPENDIX

Model Physics

Tables A1 and A2 give a summary of the variables and physical equations used in the models. Heat flux into the ground is calculated as in Brutsaert (1984), except that the KNMI model takes into account the freezing and melting energies and the FMI model has its own parameter for traffic-caused heating. This parameter has values of 10 W m^{-2} during daytime traffic (0400–1900 UTC) and 5 W m^{-2} during nighttime traffic (1900–0400 UTC). The FMI model uses a simpler approach to take into account the energy needed to melt ice and snow compared with the KNMI model. In the FMI model the surface temperature remains at 0.25°C when melting occurs instead of taking it into account in the flux calculation. The remaining energy is used to warm up the surface after all the snow and ice has melted.

Net radiation is also calculated as in Brutsaert (1984) in both models, except the KNMI model takes into account the sky-view factor (0.9). The use of the sky-view factor reduces the amount of longwave radiation from the atmosphere but takes into account the longwave radiation emitted from the surroundings. In the initialization both models calculate a correction factor for either long- or shortwave radiation. This also has an effect on the net radiation, which is explained in more detail in section 2a. Figure A1 shows the average net radiation, sensible heat, and latent heat fluxes in the 0300 UTC model runs. Other model runs also identify similar behavior. In general, the KNMI model has more positive net radiation than the FMI model.

The boundary layer conductance and stability parameters are calculated using an iterative procedure in both models. Although the equations for these parameters are rather different, the results with the same input values produce boundary layer conductance values of similar quantity when tested with $T_s = 5.0$°C, $T_a = 0.0$°C, $z_m = 0.001$ m, $z_t = 0.001$ m, and varying the wind speed from 1 to 11.5 m s^{-1}. However, the FMI model uses a larger roughness length for momentum ($z_m = 0.4$ m),

TABLE A1. Definitions of variables used in Table A2.

Variable	Definition
G	Heat flux into the ground
I_{net}	Net radiation at the surface
H	Sensible heat flux
LE	Latent heat flux
PC	Heat flow due to melting
PF	Heat flow due to freezing
Tr	Heating caused by traffic
α_s	Surface albedo
I_g	Downwelling shortwave radiation
ε_l	Absorption of longwave radiation
I_L	Downwelling longwave radiation
ε_s	Surface emittance
σ	Stefan–Boltzmann constant
T_s	Surface temperature
BLC	Boundary layer conductance
θ_s	Potential temperature at surface
θ_a	Potential temperature at 2 m
T_a	Air temperature
c_a	Specific heat of air
W_s	Water on the road
W_{max}	Maximum amount of water on the road
ρ_a	Density of air
k	von Kármán's constant
u^*	Friction velocity
z_T	Temperature measurement height
d	Zero-plane displacement
z_h	Roughness length for heat
Ψ_h	Stability correction factor for heat
Ψ_{h0}	Stability factor for heat at the height of z_h
z_W	Wind speed measurement height
z_m	Roughness length for momentum
Ψ_m	Stability factor for momentum
Ψ_{m0}	Stability factor for momentum at the height of z_m
ζ	Stability parameter
g	Gravitational constant
L_v	Specific heat of evaporation
q_a	Specific humidity for air at 2 m
q_s	Specific humidity for air at surface
ρ_m	Density of moist air
γ	Psychometric constant
e_s	Water vapor pressure of the surface
e_a	Water vapor pressure of the air
r_o	Aerodynamic resistance
T	Ground temperature
z	Vertical distance in the ground
t	Time
κ_g	Thermal diffusivity
k_g	Heat conductivity
ρ_g	Ground density
c_g	Specific heat capacity of the ground
T_{clim}	Climatological temperature deep in the ground (10.0°C)
A	0.6
ω	$2\pi/365$
z_b	Depth of the bottom layer
z_d	2.7 m

which leads to much stronger coupling of the road to the atmosphere. Consequently, the absolute sensible heat flux values are larger in the FMI model than in the KNMI model in general. Another reason for this behavior is that the temperature of the uppermost layer in the FMI model rises much faster than the surface-layer temperature in the KNMI model, which causes a larger temperature difference between the surface and the air. The difference is not that great in the verification scores because the output surface temperature in the FMI model is given as average of the top two layers. This average temperature is also used when stability parameters and the boundary layer conductance are calculated.

The equations for latent heat flux are also quite different in the two models. In general, the absolute values of the latent heat flux are greater in the FMI model than in the KNMI model. The main reason for this is again the larger roughness length in the FMI model. With similar input values and wind speeds greater than $1\,\text{m}\,\text{s}^{-1}$, the equations give latent heat flux values that are closer to each other when tested with $T_s = 5.0$°C, $T_a = 3.0$°C, $z_m = 0.001$ m, $z_t = 0.001$ m, Rh = 50%, $W_s = 0.1$ mm, and varying the wind speed from 1.5 to $11.5\,\text{m}\,\text{s}^{-1}$. However, the FMI model equations tend to still give larger absolute values. The FMI model also allows thicker layers of water and ice on the surface; so, more energy is required for evaporation. In the FMI model, the maximum limit for water storage is 2 mm, for snow it is 100 mm, for ice it is 20 mm, and for frost it is 2 mm. The values are given in water equivalent millimeters. In the KNMI model the maximum storage values for water and ice are 0.2 mm. The value for the psychrometric constant γ in the FMI model has been developed using values from Calder (1990). The value for aerodynamic resistance r_o in the FMI model is determined by a modified version of the equation given by Tourula and Heikinheimo (1998). Restrictions for low wind speeds in the FMI model are used because the divider in the equation becomes small with low wind speeds and gives quite large values for aerodynamic resistance. With wind speeds of $1\,\text{m}\,\text{s}^{-1}$ and with the other input values mentioned above, the FMI model equation gives a much larger absolute latent heat flux value than the KNMI model equation because of the usage of a constant value of $30\,\text{sm}^{-1}$ for the aerodynamic resistance.

Heat transfer in the ground is calculated in the same way in both models except the FMI models uses a different solving method for the differential equation in the initialization phase. In this phase the FMI model uses an algorithm obtained by solving the heat transfer equation by a time-centered Crank–Nicholson scheme, and the

TABLE A2. Summary of the main physical equations used in the KNMI and FMI models.

	KNMI	FMI
G	$I_{net} - H - \mathrm{LE} + \mathrm{PC} + \mathrm{PF}$ (Brutsaert 1984)	$I_{net} - H - \mathrm{LE} + \mathrm{Tr}$ (Brutsaert 1984)
I_{net}	$(1-\alpha_s)I_g + 0.9\varepsilon_l I_L + 0.1\varepsilon_s\sigma T_a^4 - \varepsilon_s\sigma T_s^4$ (Brutsaert 1984)	$(1-\alpha_s)I_g + \varepsilon_l I_L - \varepsilon_s\sigma T_s^4$ (Brutsaert 1984)
H	$\mathrm{BLC}(\theta_s - \theta_a)$ (Garratt 1992)	$\mathrm{BLC}(T_s - T_a)$ (Campbell 1985, 1986)
BLC	$\dfrac{c_a\rho_a k^2 u^*}{\left[\ln\left(\frac{z_W}{z_m}\right) - \Psi_m + \Psi_{m0}\right]\left[\ln\left(\frac{z_T}{z_h}\right) - \Psi_h + \Psi_{h0}\right]}$ (De Bruin 1994)	$\dfrac{c_a\rho_a k u^*}{\ln\left(\frac{z_T - d + z_h}{z_h}\right) + \Psi_h}$ (Campbell 1985, 1986)
u^*	$\dfrac{ku}{\ln\left(\frac{z_W}{z_m}\right) - \Psi_m + \Psi_{m0}}$ (Van Ulden and Holtslag 1985)	$\dfrac{ku}{\ln\left(\frac{z_W - d + z_m}{z_m}\right) + \Psi_m}$ (Campbell 1985, 1986)
Ψ_h, Ψ_m stable conditions	-5ζ (Van Ulden and Holtslag 1985)	4.7ζ (Campbell 1985, 1986)
Ψ_h, Ψ_m unstable conditions	$x = (1-16\zeta)^{1/4}$ $\Psi_h = 2\ln\left(\frac{1+x^2}{2}\right)$ $\Psi_m = 2\ln\left(\frac{1+x}{2}\right) + \ln\left(\frac{1+x^2}{2}\right) - 2\arctan(x) + \pi/2$ (Van Ulden and Holtslag 1985)	$\Psi_h = -2\ln\left(\frac{1+\sqrt{1-16\zeta}}{2}\right)$ $\Psi_m = 0.6\Psi_h$ (Campbell 1985, 1986)
ζ	$-\dfrac{kz_T gH}{c_a\rho_a T_a u^{*3}}$ (Campbell 1985, 1986)	$-\dfrac{kz_T gH}{c_a\rho_a T_a u^{*3}}$ (Campbell 1985, 1986)
LE	$\dfrac{-\rho_a k^2 L_v u(q_a - q_s)(W_s/W_{max})^{1/2}}{\left[\ln\left(\frac{z_W}{z_m}\right) - \Psi_m + \Psi_{m0}\right]\left[\ln\left(\frac{z_T}{z_h}\right) - \Psi_h + \Psi_{h0}\right]}$ (De Bruin 1994)	$\dfrac{\rho_m c_a}{\gamma}\dfrac{e_s - e_a}{r_o}$ (Calder 1990)
γ		$0.1 \times (0.00063T_a + 0.47496)$
r_o, $u > 1\ \mathrm{m\ s^{-1}}$		$\dfrac{\left[\ln\left(\frac{z_W + z_m}{z_m}\right) + \Psi_m\right]\left[\ln\left(\frac{z_W + z_h}{z_h}\right) + \Psi_h\right]}{k^2 u}$ (Tourula and Heikinheimo 1998)
r_o, $u \le 1\ \mathrm{m\ s^{-1}}$		$30\ \mathrm{s\,m^{-1}}$
Heat transfer in the ground	$\dfrac{\delta T(z,t)}{\delta t} = \kappa_g\dfrac{\delta^2 T(z,t)}{\delta z^2}$, $\kappa_g = \dfrac{k_g}{\rho_g c_g}$ (Garratt 1992)	$\dfrac{\delta T(z,t)}{\delta t} = \kappa_g\dfrac{\delta^2 T(z,t)}{\delta z^2}$, $\kappa_g = \dfrac{k_g}{\rho_g c_g}$ (Garratt 1992)
Temp of the bottom layer	Free boundary condition for normal road Exchange with 2-m temperature for bridges	$T_{clim} + A\sin\left(\omega t - \dfrac{z_b}{z_d}\right)$ (Campbell 1985, 1986)
Numerical solving method for heat transfer	Explicit scheme, forward Euler method	Initialization phase before coupling: time-centered Crank–Nicholson scheme used to solve equation numerically; Thomas algorithm (Campbell 1985) used to solve the obtained tridiagonal matrix system Coupling phase and forecast phase: explicit scheme, forward Euler method

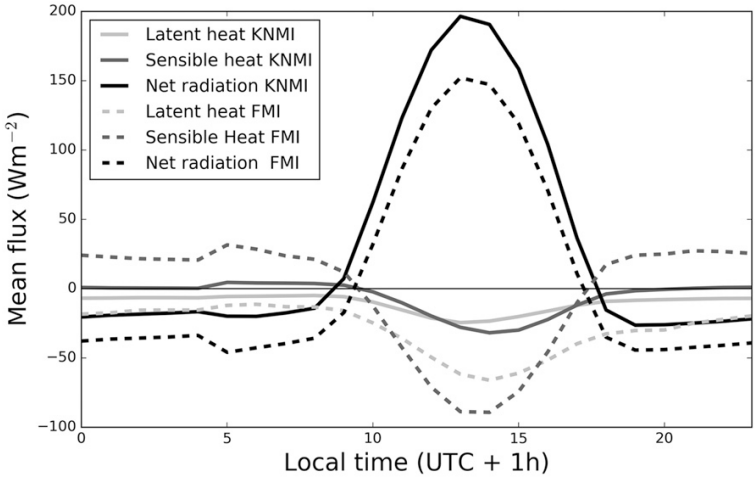

FIG. A1. The averages of net radiation (black lines), latent heat flux (light gray lines), and sensible heat flux (dark gray lines) in the 0300 UTC model runs as a function of local time. The values for the KNMI model are shown by continuous lines and the values for FMI model by dashed lines.

resulted tridiagonal matrix system is solved iteratively by the Thomas algorithm (Campbell 1985). As the lower boundary conditions, the model uses a climatological average that changes depending on the time of the year. The model was modified to use a simpler forward Euler method as the numerical solution to the heat transfer when coupling was added to the model. This method is used during the coupling phase and onward because the coupling did not work well with the Thomas algorithm–based solving method. The time step is also changed in the FMI model when the coupling phase starts. Before this change is made, the FMI model uses a time step of 5 min in the initialization, but afterward the time step is reduced to 30 s. The KNMI model implements the forward Euler method during the whole model run with a time step of 10 s and the bottom-layer temperature can evolve freely. On bridges the heat transfer from the air below also affects the bottom-layer temperature.

REFERENCES

Baas, P., and H. Van den Brink, 2014: The added value of the high-resolution HARMONIE model for deriving the HBCs: Contribution to WP 1 of the WTI 2017 Wind Modelling Project. KNMI Final Rep., 54 pp. [Available online at publicaties.minienm.nl/download-bijlage/61534/report-dec13-added-value-20140207-final.pdf.]

Bengtsson, L., and Coauthors, 2017: The HARMONIE-AROME model configuration in the ALADIN-HIRLAM NWP system. *Mon. Wea. Rev.*, doi:10.1175/MWR-D-16-0417.1, in press.

Bogren, J., T. Gustavsson, M. Karlsson, and U. Postgård, 2000: The impact of screening on road surface temperature. *Meteor. Appl.*, **7**, 97–104, doi:10.1017/S135048270000150X.

Brutsaert, W., 1984: *Evaporation into the Atmosphere: Theory, History, and Applications*. D. Reidel, 299 pp.

Calder, I. R., 1990: *Evaporation in the Uplands*. John Wiley and Sons, 148 pp.

Campbell, G. S., 1985: *Soil Physics with BASIC*. Elsevier, 150 pp.

——, 1986: *An Introduction to Environmental Biophysics*. Springer-Verlag, 159 pp.

Chapman, L., and J. E. Thornes, 2011: What spatial resolution do we need for a route-based road weather decision support system? *Theor. Appl. Climatol.*, **104**, 551–559, doi:10.1007/s00704-011-0433-9.

——, ——, and A. V. Bradley, 2001: Modeling of road surface temperature from a geographical parameter database. Part 2: Numerical. *Meteor. Appl.*, **8**, 421–436, doi:10.1017/S1350482701004042.

Crevier, L. P., and Y. Delage, 2001: A new model for road-condition forecasting in Canada. *J. Appl. Meteor.*, **40**, 2026–2037, doi:10.1175/1520-0450(2001)040<2026:MANMFR>2.0.CO;2.

David, Y., and M. L. Portal, Eds., 1985: Electronic traffic aids on major roads. Commission of the European Communities Rep. 9835, Science and Technology Policy Series, 68 pp. [Available online at https://bookshop.europa.eu/en/electronic-traffic-aids-on-major-roads-cost-30-bis-pbCDNW85003/.]

De Bruin, H., 1994: *Micrometeorologie*. Landbouwuniversiteit Wageningen, 159 pp.

Ebert, E., and Coauthors, 2013: Progress and challenges in forecast verification. *Meteor. Appl.*, **20**, 130–139, doi:10.1002/met.1392.

Efron, B., and R. J. Tibshirani, 1993: *An Introduction to the Bootstrap*. Chapman and Hall, 436 pp.

Fujimoto, A., A. Saida, and T. Fukuhara, 2012: A new approach to modeling vehicle-induced heat and its thermal effects on road surface temperature. *J. Appl. Meteor. Climatol.*, **51**, 1980–1993, doi:10.1175/JAMC-D-11-0156.1.

Garratt, J. R., 1992: *The Atmospheric Boundary Layer*. Cambridge University Press, 316 pp.

Hogan, R. J., and I. B. Mason, 2011: Deterministic forecasts of binary events. *Forecast Verification: A Practitioner's Guide in*

Atmospheric Science, 2nd ed., I. T. Jolliffe and D. B. Stephenson, Eds., John Wiley and Sons, 31–59, doi:10.1002/9781119960003.ch3.

Jacobs, W., and W. E. Raatz, 1996: Forecasting road-surface temperatures for different site characteristics. *Meteor. Appl.*, **3**, 243–256, doi:10.1002/met.5060030306.

Juga, I., P. Nurmi, and M. Hippi, 2013: Statistical modelling of wintertime road surface friction. *Meteor. Appl.*, **20**, 318–329, doi:10.1002/met.1285.

Kangas, M., M. Heikinheimo, and M. Hippi, 2015: RoadSurf—A modelling system for predicting road weather and road surface conditions. *Meteor. Appl.*, **22**, 544–533, doi:10.1002/met.1486.

Karsisto, V., P. Nurmi, M. Kangas, M. Hippi, C. Fortelius, S. Niemelä, and H. Järvinen, 2016: Improving road weather model forecasts by adjusting the radiation input. *Meteor. Appl.*, **23**, 503–513, doi:10.1002/met.1574.

Masson, V., and Coauthors, 2013: The SURFEXv7.2 land and ocean surface platform for coupled or offline simulation of Earth surface variables and fluxes. *Geosci. Model Dev.*, **6**, 929–960, doi:10.5194/gmd-6-929-2013.

Nysten, E., 1980: Determination and forecasting of road surface temperature in the COST 30 automatic road station. Finnish Meteorological Institute Tech Rep. 23, 32 pp.

Rayer, P. J., 1987: The Meteorological Office road surface temperature model. *Meteor. Mag.*, **116**, 180–191.

Roebber, P. J., 2009: Visualizing multiple measures of forecast quality. *Wea. Forecasting*, **24**, 601–608, doi:10.1175/2008WAF2222159.1.

Seity, Y., P. Brousseau, S. Malardel, G. Hello, P. Bénard, F. Bouttier, C. Lac, and V. Masson, 2011: The AROME-France convective-scale operational model. *Mon. Wea. Rev.*, **139**, 976–991, doi:10.1175/2010MWR3425.1.

Shao, J., and P. J. Lister, 1996: An automated nowcasting model of road surface temperature and state for winter road maintenance. *J. Appl. Meteor.*, **35**, 1352–1361, doi:10.1175/1520-0450(1996)035<1352:AANMOR>2.0.CO;2.

Thornes, J. E., 1984: The prediction of ice formation on motorways in Britain. Ph.D. thesis, University of London, 404 pp. [Available online at http://discovery.ucl.ac.uk/1348922/1/336707.pdf.]

——, and J. Shao, 1991: A comparison of UK ice prediction models. *Meteor. Mag.*, **118**, 93–99.

Tourula, T., and M. Heikinheimo, 1998: Modelling evapotranspiration from a barley field over the growing season. *Agric. For. Meteor.*, **91**, 237–250, doi:10.1016/S0168-1923(98)00065-3.

Undén, P., and Coauthors, 2002: HIRLAM-5 scientific documentation. HIRLAM-5 Project, 144 pp. [Available online at http://hirlam.org/index.php/hirlam-documentation/doc_download/270-hirlam-scientific-documentation-december-2002.]

Van Ulden, A. P., and A. A. M. Holtslag, 1985: Estimation of atmospheric boundary layer parameters for diffusion applications. *J. Climate Appl. Meteor.*, **24**, 1196–1207, doi:10.1175/1520-0450(1985)024<1196:EOABLP>2.0.CO;2.

Venäläinen, A., and M. Kangas, 2003: Estimation of winter road maintenance costs using climate data. *Meteor. Appl.*, **10**, 69–73, doi:10.1017/S1350482703005073.

WMO, 2014: Forecast verification for the African severe weather forecasting demonstration projects. WMO 1132, 31 pp. [Available at https://www.wmo.int/pages/prog/www/Documents/1132_en.pdf.]

Yang, C. H., D. G. Yun, and J. G. Sung, 2012: Validation of road surface temperature prediction model using real-time weather forecasts. *KSCE J. Civ. Eng.*, **16**, 1289–1294, doi:10.1007/s12205-012-1649-7.

PLoS ONE, 14(2): e0211702,
doi: 10.1371/journal.pone.0211702

III

PLOS ONE

RESEARCH ARTICLE

Mobile road weather sensor calibration by sensor fusion and linear mixed models

Lauri Lovén[1]*, **Virve Karsisto**[2], **Heikki Järvinen**[3], **Mikko J. Sillanpää**[1], **Teemu Leppänen**[1], **Ella Peltonen**[1], **Susanna Pirttikangas**[1], **Jukka Riekki**[1]

1 Infotech Oulu, University of Oulu, Oulu, Finland, **2** Finnish Meteorological Institute, Helsinki, Finland, **3** University of Helsinki, Helsinki, Finland

* lauri.loven@oulu.fi

Citation: Lovén L, Karsisto V, Järvinen H, Sillanpää MJ, Leppänen T, Peltonen E, et al. (2019) Mobile road weather sensor calibration by sensor fusion and linear mixed models. PLoS ONE 14(2): e0211702. https://doi.org/10.1371/journal.pone.0211702

Editor: Jie Zhang, Newcastle University, UNITED KINGDOM

Received: July 24, 2018

Accepted: January 19, 2019

Published: February 7, 2019

Data Availability Statement: All relevant data files are available from the OSF database (project name "Surface temperature data fusion", osf.io/32aqf).

Funding: The 6Genesis Flagship program is funded by the Academy of Finland (grant 318927). The program funding was awarded to the University of Oulu. The AI Enhanced Mobile Edge Computing project is funded by the Future Makers program of Jane and Aatos Erkko Foundation and the Technology Industries of Finland Centennial Foundation. The project funding was awarded to

Abstract

Mobile, vehicle-installed road weather sensors are becoming ubiquitous. While mobile sensors are often capable of making observations on a high frequency, their reliability and accuracy may vary. Large-scale road weather observation and forecasting are still mostly based on stationary road weather stations (RWS). Though expensive, sparsely located and making observations on a relatively low frequency, RWS' reliability and accuracy are well-known and accommodated for in the road weather forecasting models. Statistical analysis revealed that road weather conditions indeed have a great effect on how the observations of mobile and stationary road weather temperature sensors differ from each other. Consequently, we calibrated the observations of mobile sensors with a linear mixed model. The mixed model was fitted fusing ca. 20 000 pairs of mobile and RWS observations of the same location at the same time, following a rendezvous model of sensor calibration. The calibration nearly halved the MSE between the observations of the mobile and the RWS sensor types. Computationally very light, the calibration can be embedded directly in the sensors.

Introduction

Mobile, vehicle-installed sensors and road weather station (RWS) networks can together provide denser and higher quality information than either alone. They can support optimization of maintenance operations, such as snow clearance and prevention of slipperiness, and generation of real-time warnings for road users. Accurate now-casting and forecasting are keys to safe and economic operations, especially in the northern latitudes where driving conditions can vary a lot in space and time, increasing risk for accidents.

Improved technologies and increased availability of mobile observations can drastically improve the coverage and quality of observations on roads. The amount of available mobile observations have recently considerably increased. During November 2016—March 2017, for instance, vehicles fitted with Teconer Oy's optical sensors (RCM411 and RTS411) covered globally approximately 200 000 km of roads per month observing friction, surface water deposits, and road surface temperature [1].

the University of Oulu. The Intelligent Arctic trucks project is funded by Regional Council of Lapland1 and European Regional Development Fund of the European Union2. The project funding was awarded to the Finnish Meteorological Institute. The Sod5G project is funded by Regional Council of Lapland and European Regional Development Fund of the European Union. The project funding was awarded to the Finnish Meteorological Institute. TheWiRMa (Industrial Internet Applications in Winter Road Maintenance) project is funded by the Regional Council of Lapland and the Interreg Nord fund of the European Union (http://www.interregnord.com/). The fund was awarded to the Finnish Meteorological Institute. The 5G-Safe project is funded by Business Finland (https://www.businessfinland.fi/en/). The fund was awarded to the Finnish Meteorological institute. Finally, V. Karsisto received a grant for PhD studies related to "Using car observations in road weather forecasting." The grant was awarded by the Finnish Cultural foundation (https://skr.fi/en). Grant number is 00150364. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

There have been several studies about the usability of mobile observations. For example, mobile road condition monitoring was tested in Finland already in 1998—1999 [2]. Also Stern et al. [3] studied the reliability of mobile observations, while Koller et al. [4] examined the quality of the observations with the native sensors of a vehicle. In 2013, Finnish transport agency compared optical friction and temperature meters [5]. The study compared, among other things, Teconer's optical road surface temperature sensor RTS411 with RWS measurements. RTS411 gave typically 1.2˚C warmer values than the RWS. Such a large difference between observations calls for sensor calibration.

Sensor calibration unifies sensor data that one or multiple sensors collect for a particular application. Drifting, zero offset errors, or variations in the manufacturing process can cause even sensors by the same manufacturer to yield different readings in the same conditions, with systematic or random errors. Measurement errors are aggravated when the sensors, during operation or storage, are subjected to varying environmental conditions such as light, temperature, humidity, hysteresis or shock. In addition, the selected sensor technology may initially provide low signal to noise-ratio, which makes repeatable measurements challenging. Thus, re-calibration in the field may be needed after initial factory calibration. This problem is magnified when different types of the same sensor modality are used to measure the same physical phenomena. In such a case, multiple sensor readings may be based on measuring completely different, but related, parameters of the monitored phenomena.

Typically, single sensors are calibrated with some physical reference, e.g. gravity, that provides a direct mapping of measured values to standardized units. For linear sensors, simple calibration can be based on reference points. For non-linear sensors, calibration typically requires multi-point curve fitting methods. For example, air quality sensor applications [6] use highly accurate weather stations as reference points, providing the ground truth. However, the sensing range and frequency provided by such stations is limited and the distances between stations result in a low spatial resolution.

The coverage and resolution of sparse, stationary sensors can be improved with mobile devices with integrated sensors, e.g. vehicle-installed smartphones [7]. However, smartphone sensing devices suffer from noisy measurements due to low-cost sensor technology [6]. Further, inexperienced users can cause errors in the data, for example by installing the sensors incorrectly. Finally, manual calibration by users may introduce uncertainty. Automatic calibration methods are thus used to eliminate the cumbersome and error-prone manual calibration [8]. One such automatic calibration method is the rendezvous model [9–12]. In the rendezvous model, observations by two or more sensors, mobile or stationary, are collected when the sensors are co-located, i.e. at the same time in the same place. The mobile sensors are calibrated by comparing the rendezvous observations. In addition to ensuring the spatio-temporal identity of the observations, the locations of the static reference points need to be carefully considered [8].

Sensor fusion refers to merging information from two or more sensors, located spatially close to each other. Sensor fusion helps with problems related to low spatial resolution and unreliable users. It has previously been used for example for autonomous vehicle navigation, improving lane [13] and road potholes [14] recognition. Rendezvous model uses sensor fusion to unify data from different sensors.

This study proposes a sensor fusion based method to calibrate mobile road weather sensors. Specifically, road surface temperature sensors are calibrated with the rendezvous model. The model compares spatio-temporally co-located observations by sparse Vaisala RWS sensors with the dense but possibly less reliable mobile observations provided by vehicle-installed Teconer sensors RCM411 and RTS411. The aim is to first chart the statistical characteristics of

the mobile observations when compared to the RWS observations, and then calibrate the mobile sensors to agree with the RWSs.

The paper is organized as follows: Section *Methods and Data* presents the methodology, data, and statistical models used in the study, Section *Results and discussion* presents the results and discusses the outcome, and finally Section *Conclusion* concludes the study.

Methods and data

Overview

This study presents a novel, sensor fusion based method to calibrate mobile surface temperature sensors to agree with RWSs. The methodology is summarized as follows:

1. This study uses RWS and mobile data, collected during winter seasons between 2014–2017.
2. Rendezvous, i.e. occasions where the mobile sensors have passed one of the RWSs, are identified. The data on these rendezvous comprises the full data set, while the remaining data is discarded.
3. Relationships between the mobile observations and the RWS measurements are analysed with statistical models. Two particular research questions are considered:
   a. How are the mobile observations related to environmental conditions such as weather?
   b. How can we calibrate the mobile sensors, i.e. improve the accuracy of the mobile observations when compared to the RWS observations?
   Statistical inference models are built to answer the first question. Statistical prediction models are built to answer the second question.

Data

This study used road weather data collected by 251 stationary Vaisala RWSs, and 27 vehicles installed with Teconer RCM411/RTS411 sensors. Fig 1 shows the locations of the RWSs, mostly deployed at the major roads around southern and central Finland. The full data set contains 25944 observations. However, few RWSs were fitted with a full array of sensors, so in most cases some of the observed variables are missing.

The RWSs measure multiple weather variables, such as temperature and wind conditions. We used a representative subset of the available variables fitting to the road-condition analysis, which are listed in Table 1. First, road surface temperature is measured at most of the stations by an asphalt embedded DRS511 sensor [15]. Further, the DRS511 sensor also estimates overall road status. This can be, for example, "wet", "icy", or "snowy". Observations whose road status is "error" are discarded.

Some RWSs have optical Vaisala DST111 sensors [16], measuring road surface temperature by the infrared radiation emitted by the road surface. DST111 measures also air temperature. However, optical sensors at Finnish RWSs are more commonly Vaisala DSC111s [17], which detect water, ice, and snow deposits on the road. The method is based on absorption wavelengths of water and ice. The device transmits infrared radiation with certain wavelengths, and the deposits are determined from the radiation backscattered from the surface [18]. For Vaisala DSC111, the resolution for water, ice and snow is reported as 0.01 mm [17]. DSC111 determines also the road status ("wet", "icy" etc). (Observations with road status "error" were discarded.) DSC111 and DST111 sensors are typically installed on poles beside the road.

Some of the road weather stations contained multiple surface temperature sensors. These sensors were either optical or installed in the asphalt. In this study, only one such sensor for a

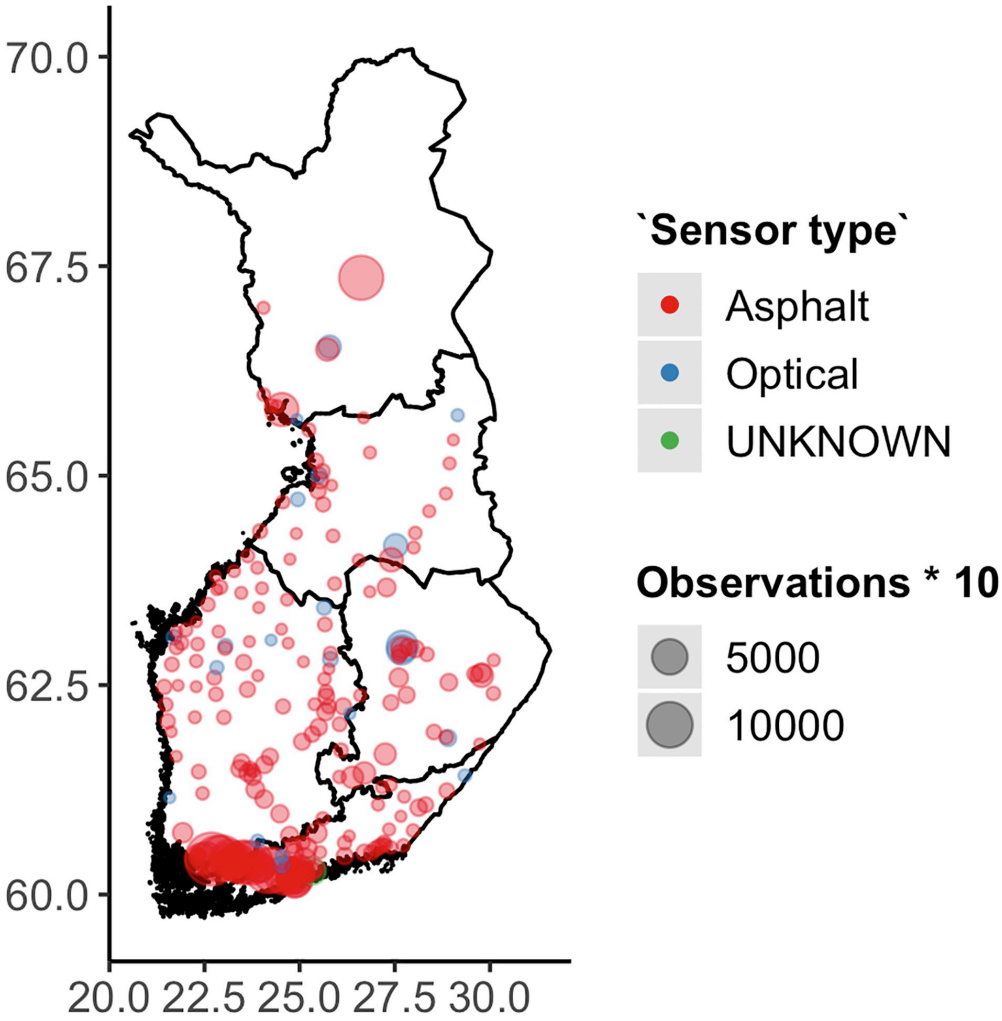

Fig 1. Spatial distribution of data. Coordinates are in latitude and longitude.

https://doi.org/10.1371/journal.pone.0211702.g001

station was used at a time. Thus, the Vaisala sensor identification code, telling apart the individual sensor devices, also differentiates between the RWSs.

The optical Teconer RCM411 road condition sensor, fitted on vehicles, has the same operational principle as the DSC111. It observes, among other things, overall road status (e.g. "wet", "icy" etc.) [1]. An add-on on the RCM411, a Teconer RTS411 sensor measures road surface temperature based on infrared radiation similarly as Vaisala DST111.

RCM411+RTS411 device pairs can be connected to a nearby cell phone by a Bluetooth connection. The cell phone is used to obtain the location, direction and speed of the vehicle. These are also included in Table 1.

Table 1. Data description.

Type	Source	Variable	Notes
Mobile	Teconer RTS411	Road surface temperature	Continuous (˚C).
	Teconer RCM411	Road state	Categorical: Dry, Moist, Wet, Slush, Ice, or Snow/Frost.
	Teconer RCM411+GPS	Location	Continuous (lat,lon).
	Teconer RCM411	Sensor ID code	Categorical.
RWS	Vaisala DRS511/DST111	Road surface temperature	Continuous (˚C).
	Vaisala DST111	Air temperature	Continuous (˚C).
	Vaisala DRS511+DST111	Road state	Categorical: Dry, Moist, Wet, Wet&Salty, Frost, Snow, Ice, Moist&Salty, Slushy
	Vaisala DSC111	Road state	(as above)
	Vaisala DSC111	Water deposits	Continuous (mm).
	Vaisala DSC111	Ice deposits	Continuous (mm).
	Vaisala DSC111	Snow deposits	Continuous (mm).
	Vaisala DRS511/DST111	Sensor type	Categorical: Optical, Asphalt, or Unknown.
	Vaisala DRS511/DST111	Sensor ID code	Categorical. Individual id code for each vehicle installation.
Circumstantial		Year	Continuous.
		Season	Categorical: Fall (Sep–Nov), Winter (Dec–Mar), Spring (Apr).

https://doi.org/10.1371/journal.pone.0211702.t001

Teconer device identification codes differentiate between individual RCM411+RTS411 device pairs. However, as one device pair could have multiple vehicle installations over the observation period, and each installation could have a different calibration level, we modified the identification codes such that each installation had its own identification code.

The study data contains observations from three cold periods (September—April) between years 2014 and 2017. Firstly, we compare the mobile observations to the RWS locations, and then select those data points where a mobile sensor passed by an RWS within 50m or less. Each such occasion constitutes a rendezvous. As the mobile sensor observation rate is once per second, one rendezvous might comprise several observations. Thus, the numerical observations are averaged, and for the non-numerical observations (e.g. road status), we choose the one with the smallest distance between the RWS and the mobile sensors. However, the non-numerical observations stay constant during the vast majority of rendezvous', indicating our choice of 50m radius is well chosen.

Rendezvous time is considered to be the middle time point between the first and the last observation within one pass. The RWS observation rate is once per 5-10 minutes, depending on the station. Thus, there is often a time discrepancy between an RWS observation and the mobile sensor pass time. The RWS observation with the smallest absolute time difference to the pass time is used in this study.

Explorative analysis

Surface temperature densities for each RWS road status are depicted in Fig 2. They are almost identical, which indicates that the mobile and the RWS observations largely agree. However, the mobile observations appear slightly higher than the RWS observations. A scatterplot (see Fig 3) between the mobile and the RWS surface temperature observations further reveals that while the observations largely agree, following a linear relationship, between 0˚C and 10˚C the LOESS prediction line for the asphalt embedded sensor takes a sudden downward turn.

Fig 4 shows that the correlation between RWS air and surface temperatures is strong. Again, there is a change in slope between 0˚C and 10˚C for the asphalt embedded sensor.

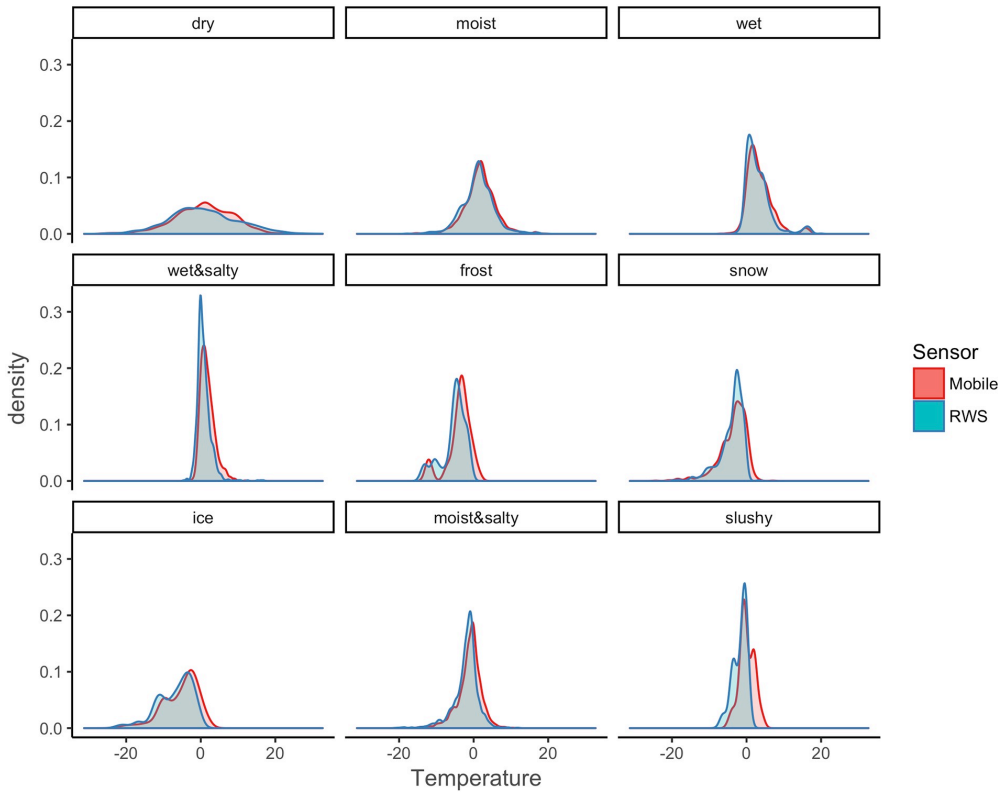

Fig 2. Surface temperature observations for each RWS road status.

https://doi.org/10.1371/journal.pone.0211702.g002

The number of observations on individual RWS and mobile sensors follow similar patterns (Fig 5). There are few sensors with hundreds or even thousands of observations, and a long tail with just a few.

Average differences between individual RWS and mobile sensors vary largely (Fig 6). The differences seem to depend more on the mobile sensor device than the RWS. One mobile device tends to diverge from the RWSs roughly the same amount, regardless of the RWS. This suggests that the vehicle and the installation of the mobile sensor device affect the observations significantly.

The number of distinct sensors per year is listed in Table 2. The number of mobile sensor installations are steadily increasing, and their reach of RWSs consequently rising. (Note that the last observation was taken at 2017-04-30).

Categorizing the observations of the full data set by both the mobile and the RWS weather status observations, the data are distributed as depicted in Table 3. While the diagonal of Table 3 dominates, there are many observations where the mobile and the RWS weather states differ on a fundamental level. In part, this is explained by the way our observations were collected, as the driving conditions can be different within the at most 50m distance between the

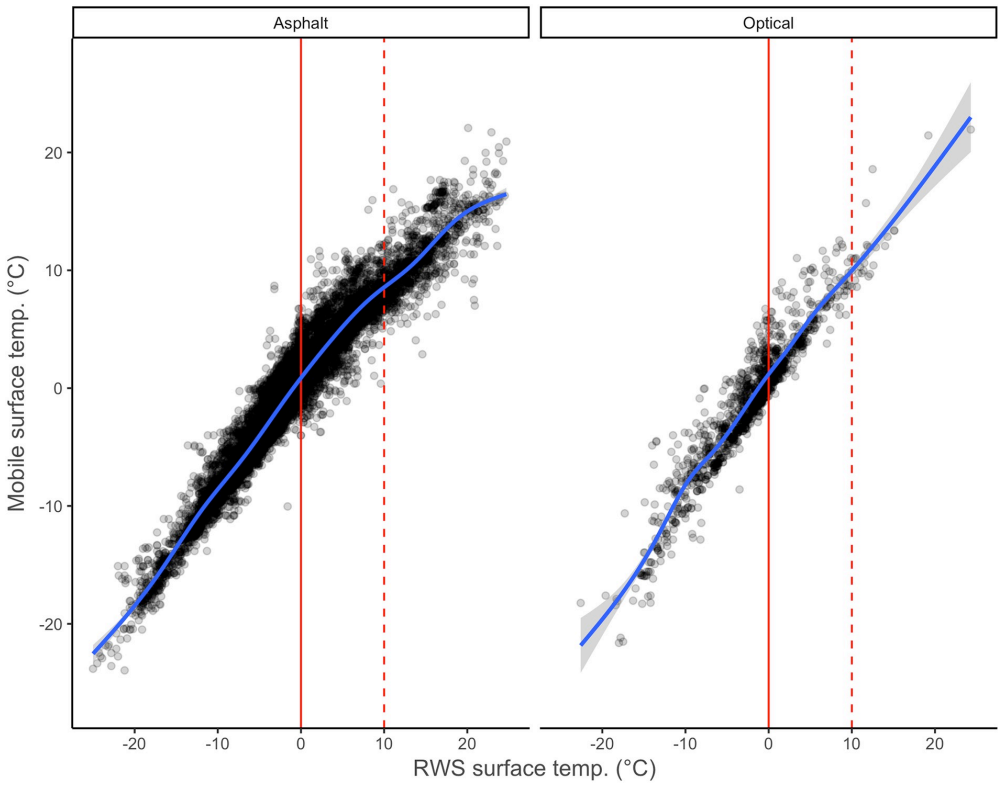

Fig 3. Scatterplot of RWS and mobile surface temperature observations. Nonparametric (LOESS) prediction line is shown in blue, the 0˚C level in solid red, and 10˚C in dashed red.

https://doi.org/10.1371/journal.pone.0211702.g003

mobile and the RWS sensor. Still, it is clear that the devices interpret their surrounding environments in different ways.

Statistical models

A linear regression model estimates a linear and additive relationship between independent variables and a response. Such a model may contain a varying number of fixed effects, estimating the *ceterus paribus* relationship between individual independent variables and the response. However, there may be only one random effect: the noise term.

In contrast, a linear mixed effects model [19, 20] may contain several random effects. If an independent variable is related to a random effect instead of a fixed one, the model no longer estimates a coefficient for inferring the relationship between the variable and the response. Instead, the effect is assumed to follow a zero-mean Gaussian distribution.

Linear mixed models are computationally light and can be used for both statistical inference and prediction. In this study, mixed models are first used inferentially, to find out how circumstantial variables as well as the weather conditions, as measured by both the mobile sensors and the RWSs, affect the quality of the road surface measurements. Second, linear mixed

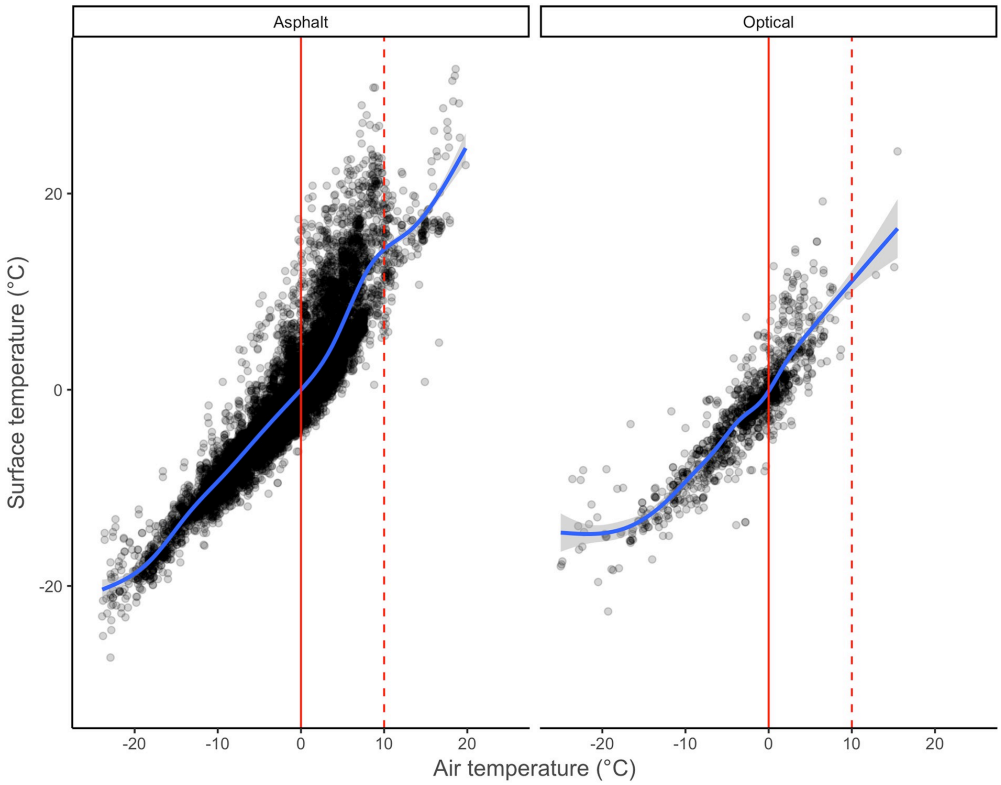

Fig 4. Scatterplot of RWS surface and air temperature observations. Nonparametric (LOESS) prediction line is shown in blue, the 0˚C level in solid red, and 10˚C in dashed red.

https://doi.org/10.1371/journal.pone.0211702.g004

models are used predictively to calibrate the mobile sensors to agree with the RWSs. A full list of variables is shown in Table 1.

Simple linear regression models provide a baseline for comparison of the inferential and the predictive cases. The baseline inference models A0 and A00 are defined as:

$$T_{s,\mathrm{Mob}} = X_{\mathrm{RWS,\ id}}\beta + \boldsymbol{\varepsilon}, \tag{1}$$

where $T_{s,\mathrm{Mob}}$ is mobile surface temperature, $X_{\mathrm{RWS,id}}$ is a $(n \times p)$ design matrix consisting of the intercept, the RWS surface temperature observations, and (for A00) dummy variables for the sensor id codes, controlling for the individual sensor factory calibration levels. Factory calibration level here refers to a device's reading when the surface temperature is actually 0˚C. This could be affected by the internal settings of the device as well as the installation on the vehicle (by way of e.g. proximity to exhaust fumes). $\boldsymbol{\varepsilon}$ is a vector of random noise with $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, I\sigma_e^2)$. A0 and A00 are used to provide a reference point for assessing the quality of the mixed models A1–A4 that were defined as

$$T_{s,\mathrm{Mob}} = X_{\mathrm{RWS}}\beta + Z_{\mathrm{id}}b + \boldsymbol{\varepsilon}, \tag{2}$$

where X_{RWS} is a $(n \times p)$ design matrix consisting of the intercept, the RWS weather

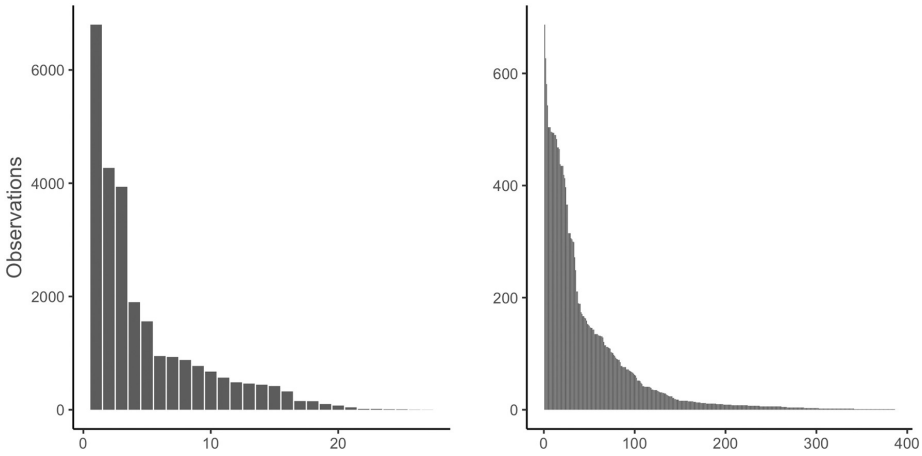

Fig 5. Histograms of observations on RWS (left) and mobile (right) sensors.

https://doi.org/10.1371/journal.pone.0211702.g005

observations, and some circumstantial data, while Z_{id} is a design matrix for the dummy variables of the mobile and RWS sensor identification codes. Finally, ε is a vector of random noise with $\varepsilon \sim \mathcal{N}(0, I\sigma_\varepsilon^2)$.

The β parameters are considered fixed effect coefficients. The parameters $b = (b_1, b_2)$ are considered random effect coefficients and correspond to the individual factory calibration levels of the devices. b_1 refers to mobile sensors and b_2 to RWS sensors, with $b_1 \sim \mathcal{N}(0, I\sigma_1^2)$ and $b_2 \sim \mathcal{N}(0, I\sigma_2^2)$. Thus, the model assumes the individual factory calibration levels of the sensor devices are normally distributed with zero means but separate variances for the two device types.

The aim of the predictive modelling is to predict the RWS surface temperature using only mobile observations. We have a baseline linear model P0:

$$T_{s,\mathrm{RWS}} = X_{\mathrm{Mob}}\beta + \varepsilon, \tag{3}$$

where $T_{s,\mathrm{RWS}}$ is RWS surface temperature, X_{Mob} is a $(n \times 2)$ design matrix consisting of the intercept and the mobile surface temperature observations, and ε is a vector of random noise with $\varepsilon \sim \mathcal{N}(0, I\sigma_\varepsilon^2)$. P0 is used to provide a reference point for assessing the prediction quality of the mixed models P1–P3, defined as:

$$T_{s,\mathrm{RWS}} = X_{\mathrm{Mob}}\beta + Z_{\mathrm{id}}b + \varepsilon, \tag{4}$$

where X_{Mob} is a $(n \times p)$ design matrix consisting of the mobile road weather observations, Z_{id} is a design matrix for the two dummy id code variables, and ε is a vector of random noise with $\varepsilon \sim \mathcal{N}(0, I\sigma_\varepsilon^2)$. Again, the β parameters are considered fixed, while the $b = (b_2, b_2)$ parameters are normally distributed.

Model quality

This study uses the conditional Akaike Information Criterion (conditional AIC) [21] to compare the quality between models. Further, mixed effect models have the following assumptions: (1) the independent variables (i.e. the columns in the model matrices X and Z) should not be correlated, (2) the residuals should have a constant variance, (3) the residuals should be

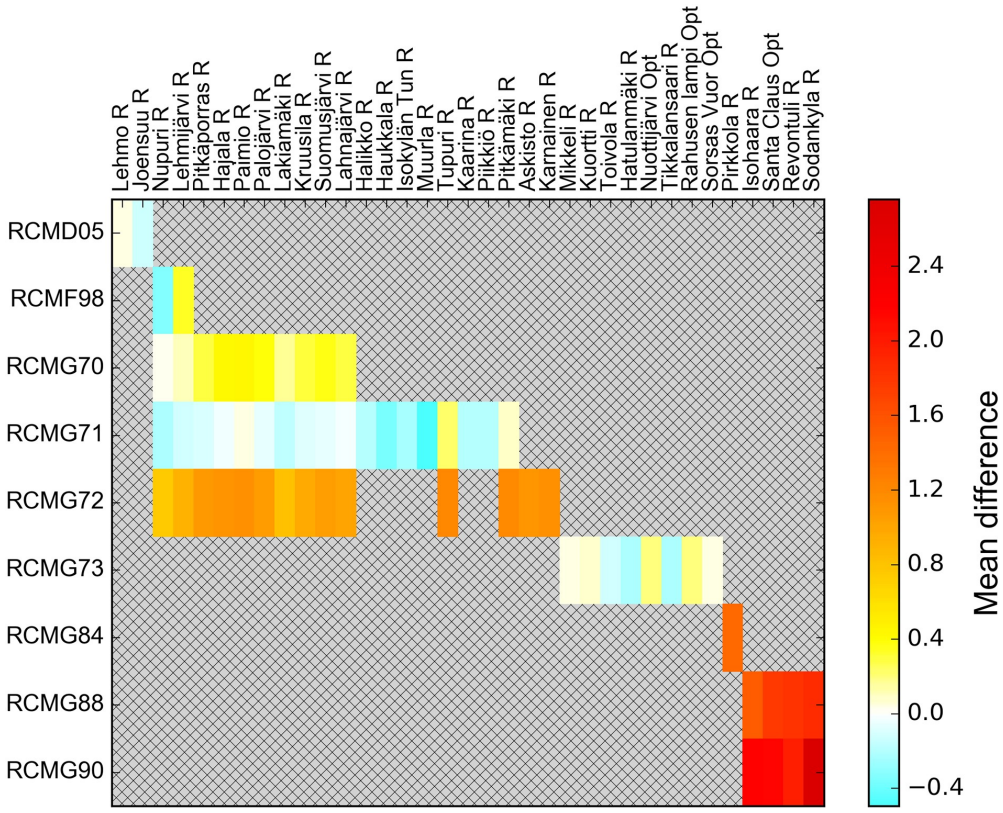

Fig 6. Average difference between each mobile–RWS observation pair. Stations marked with "R" measure surface temperature with a Vaisala DRS511, embedded in the asphalt, while stations marked with "Opt" have optical Vaisala DST111 instruments. When the RWS and/or the mobile observation is above 8˚C, the pair is discarded. Further, each mobile–RWS pair have at least 40 measurements.

https://doi.org/10.1371/journal.pone.0211702.g006

uncorrelated, and (4) normally distributed with (5) zero means, and finally, the random effects should be (6) normally distributed. The inferential model was tested for each of these assumptions.

The Variance Inflation Factors (VIFs) [22] for the inference model variables {*RWS road state, >4°C, RWS surface temp., RWS water deposits, RWS ice deposits, RWS Snow deposits, Year, Season, Distance of sensors*} are, respective, {4.4, 1.8, 2.2, 1.2, 2.0, 1.4, 1.3, 1.8, 1.0}. These are well below 10, considered as severe multicollinearity [22], and thus in agreement of Assumption 1.

Table 2. Number of distinct sensors per year.

	2015	2016	2017
Mobile	14	17	22
RWS	199	324	295

https://doi.org/10.1371/journal.pone.0211702.t002

Table 3. Number of observations by mobile and RWS road status.

Mobile	RWS								
	Dry	Moist	Wet	Wet&Salty	Frost	Snow	Ice	Moist&Salty	Slushy
Dry	4097	497	63	45	0	16	205	565	2
Moist	927	841	187	224	2	21	123	1773	2
Wet	344	1401	1363	1381	1	26	34	3113	5
Slush	10	30	39	78	0	62	31	212	12
Ice	379	160	28	42	34	648	450	582	13
Snow/frost	14	3	0	0	1	133	47	11	0

https://doi.org/10.1371/journal.pone.0211702.t003

Including the RWS air temperature variable in the model raises the RWS surface temperature VIF score above 10. This is in line with the observed high correlation in Fig 4. Consequently, RWS air temperature is excluded from the models.

S1 Fig shows the residual vs. fitted plot of the associative model A4. The plot indicates the residuals are with constant variance (Assumption 2) and with zero means (Assumption 5). Further, no autocorrelation is visible in the plot, fulfilling Assumption 3.

The Quantile-Quantile plots in S2, S3 and S4 Figs indicate validity of assumption that the residuals and random effects follow the normal distribution (assumptions 4, 6) moderately well. While the plots indicate the residual and the RWS calibration level distributions have somewhat heavier than Gaussian tails, and the mobile sensor calibration level distribution is more mesocurtic, the bulk of the standardized quantiles fall on the diagonal. Further, all distributions appear symmetrical, suggesting the divergence in the tails, while adding variance, does not add bias in our estimates.

For the predictive models, the main point of interest is the prediction accuracy. Accordingly, we measure the prediction accuracy of each model by the cross-validated mean squared error (MSE). The cross-validated MSE is calculated as the average of the MSEs of 10 unstratified folds. Further, the coefficient estimates for the predictive models are the averages of those of each fold.

Results and discussion

Inference models

Coefficient estimates and model quality indicators for the inference models are listed in Table 4, with the inference model coefficients further detailed in Fig 7.

A0 is the baseline inference model, with just the intercept and the slope adjustment to the RWS surface temperature value. A00 still has only fixed effects, but includes additionally the individual sensor calibration levels as categorical variables. A1 assumes the underlying distribution of the calibration levels is Gaussian and considers them as random effects. A2–A4 include various other variables into the model.

Model quality considerations (see section Model quality) show a slight deviation from the Gaussian assumption for the residuals, which potentially causes noise in the estimates. This is possibly one cause of the relatively large standard error seen in the model intercepts (see Table 4 and Fig 7). Apart from the underlying error term distribution, the large intercept standard error may be caused by the spatial and temporal difference between the mobile and the stationary observations.

The estimates by the baseline A0 model are well in agreement with those of the mixed A1, A3 and A4 models. On the other hand, model A00 intercept's value is markedly higher, reaching 1.2˚C. The difference between A00 and A1, a corresponding mixed model, is especially

Table 4. Estimates of fixed effect coefficients and quality metrics for the inference models.

Term	A0 (linear)	A00 (linear)	A1 (mixed)	A2 (mixed)	A3 (mixed)	A4 (mixed)
Intercept	0.53 (0.01)	1.18 (0.13)	0.54 (0.18)	1.42 (0.40)	0.61 (0.20)	0.62 (0.20)
Sensor calibr. levels	no	fixed	random	random	random	random
RWS surface temp.	0.86 (0.00)	0.87 (0.00)	0.87 (0.00)	1.05 (0.01)	0.96 (0.00)	0.96 (0.00)
> 4˚C interact.				-0.26 (0.04)	-0.29 (0.01)	-0.28 (0.01)
Moist				0.15 (0.11)	0.33 (0.03)	0.34 (0.03)
Wet				0.17 (0.14)	0.44 (0.03)	0.46 (0.03)
Wet & Salty					0.43 (0.03)	0.43 (0.03)
Snowy/frosty/slushy				0.10 (0.14)	0.19 (0.04)	0.21 (0.04)
Icy				0.13 (0.18)	0.32 (0.04)	0.33 (0.04)
Moisty & salty					0.34 (0.02)	0.34 (0.02)
Optical sensor					0.31 (0.10)	0.17 (0.10)
>4˚C interact.						1.15 (0.10)
Water deposits				-0.64 (0.34)		
Ice deposits				1.22 (0.53)		
Snow deposits				-0.26 (0.31)		
Year				0.33 (0.09)	-0.21 (0.02)	-0.21 (0.02)
Fall				-0.06 (0.16)	-0.02 (0.02)	-0.01 (0.02)
Spring				0.01 (0.12)	0.11 (0.03)§	0.09 (0.03)
conditional AIC	73138	65210	65377		59134	59004
R^2	0.927	0.951	0.953	0.954	0.965	0.966
N	20278	20278	20278	1283	20278	20278

For estimates, the standard error is provided in parentheses.

https://doi.org/10.1371/journal.pone.0211702.t004

striking, with A1 intercept at 0.54˚C. The reason behind the difference may be the uneven distributions of observations by the different sensors. As seen in Figs 5 and 6, the distribution of observations on individual sensors is very uneven, and the factory calibration levels appear to vary greatly. The simple linear model A00 is not robust enough to cope with sensors whose

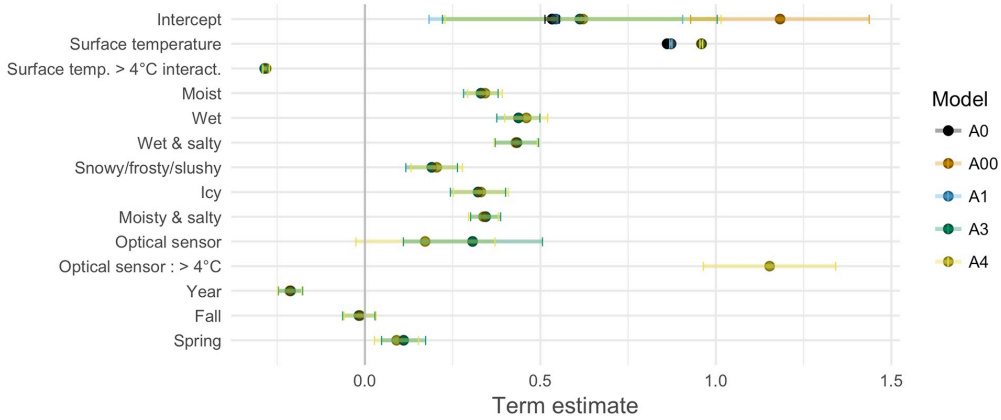

Fig 7. Estimates and 95% confidence intervals of fixed coefficients for the inference models. Model A2 excluded due to term instability.

https://doi.org/10.1371/journal.pone.0211702.g007

outlying factory calibration levels skew the intercept with even a small number of observations, which is a further justification for the use of the linear mixed models.

Interestingly, the A00 intercept exactly agrees with the 1.2˚C reported by Malmivuo [5]. Malmivuo's method investigated the medians of the mobile Teconer RTS411 and the RWS observations. The model itself is comparable with our A0, albeit somewhat more robust to outliers. The difference between the results may be due to Malmivuo's study only utilizing the measurements of one mobile sensor, whose factory calibration level, as seen in Fig 6, can be off the actual.

Models A1, A3 and A4 largely agree on their coefficient estimates, with the small confidence intervals indicating high reliability in the estimates. Ultimately, the conditional AIC model quality criteria chose model A4 as the one best describing the data.

Due to the relatively low number of observations, the mixed inference model A2 is somewhat unstable, with large confidence intervals for many coefficient estimates. However, the model suggests that ice deposits on the road tend to cause the mobile sensor device to have higher than average surface temperature readings or, conversely, the RWS observations to fall under the average. This may be caused by the asphalt embedded RWS sensors (Vaisala DRS511), whose temperature readings can vary a lot from an optical one measuring the surface of the ice.

Relying on the A4 model, in dry conditions, at 0˚C surface temperature, the mobile observations are on average 0.62 degrees higher than those of the RWS. However, there is considerable variation in the estimate, with a 95% confidence interval spanning [0.23˚C, 1.0˚C]. Further, different driving conditions have quite a strong effect on the difference. For example mobile observations are further 0.46 degrees higher in wet conditions, resulting over one degree above RWS in total.

Surface temperature coefficient is very close to 1 as expected. This indicates that mobile and RWS observations indeed almost agree when surface temperature is under 4˚C and after level adjustments provided by the different driving conditions are taken into account. A marked discrepancy appears between the observations above 4˚C, with the mobile sensor reporting observations consistently and increasingly lower than the RWS. This observation is in line with Fig 3 that shows a distinct turn in the prediction curve at around 4˚C for the asphalt embedded sensors, comprising a large majority of observations.

RWS sensor type does not have an effect on the measurements under 4˚C. However, model A4 indicates that the optical DST111 RWS sensor gives as much as 1.2˚C lower readings than the asphalt embedded DRS511 above 4˚C. This can be caused by the asphalt embedded DRS511 heating up too much in direct sunlight during sunny autumn and spring days. The observation is again in line with Fig 3, where the optical sensor has a much straighter prediction line compared to that of the one embedded in the asphalt.

The season of observation has little influence on the observations. However, there is a significant yearly trend in the measurements. Teconer observations seem to drop by ca. 0.2˚C in relation to RWS observations year-by-year. This may be caused by wear-and-tear in the mobile sensors, e.g. by dirt or scratches accumulating on the optics. Another potential cause is adjustments in the mobile sensor installations that were not included in the data available for this study. Either way, a continuous degradation calls for recurrent sensor calibration.

Prediction models

Coefficient estimates and model quality indicators for the prediction models are listed in Table 5. The mean squared error (MSE) of the predictive models P0–P3 indicate that the RWS surface temperatures predicted by model P1 follow the observed ones most closely.

Table 5. Estimates of fixed effects coefficients and performance metrics for the prediction models.

Term	Obs.	P0 (linear)	P1 (mixed)	P2 (mixed)	P3 (mixed)
Intercept		-0.62			
Teconer surface temp.		1.075	1.132	1.089	1.083
Moist interact.			-0.112		
Wet interact.			-0.093		
Snowy/frosty interact.			-0.233		
Icy interact.			-0.159		
Slushy interact.			-0.174		
Dry			-0.027	0.004	
Moist			-0.793	-0.747	
Wet			-0.958	-1.047	
Snowy/frosty			-1.001	-0.601	
Icy			-0.996	-0.508	
Slushy			-1.014	-0.946	
MSE	3.252	2.747	1.923	2.051	2.212
N	20305	20305	20305	20305	20305
folds		10	10	10	10

Estimates and metrics are both calculated by averaging the results of all folds in a 10-fold cross-validation set-up.

https://doi.org/10.1371/journal.pone.0211702.t005

Unmodified mobile observations had an MSE of 3.3. Adjusting the mobile observation with the average difference between the observations (-0.62, taken from the intercept-only regression model P0) would slightly improve the result, reaching MSE 2.7. However, the use of linear mixed models P1–P3 improve the result markedly. P1 MSE is 1.9, increasing the accuracy of the unmodified mobile observation by more than 40%.

Model P1 includes interaction terms between each road status and the RWS surface temperature, in addition to level adjustments. This changes the slope of the linear relationship between RWS and mobile observations. The following linear formula can thus be applied to calibrate the mobile Teconer RTS411 observations to better agree with the RWS:

$$T_{s,\mathrm{Adj}} = \beta_0 + \beta_1 \times T_{s,\mathrm{Mob}}, \tag{5}$$

where $T_{s,\mathrm{Adj}}$ is the adjusted mobile surface temperature, β_0 is the level coefficient and β_1 is the slope coefficient. Table 6 lists the level and slope coefficients for adjusting the mobile surface temperature in the observed road status, extracted and rounded from model P1 in Table 5.

Table 6. Linear coefficients for adjusting the mobile Teconer RTS411 road surface temperature observations.

Road status	β_0	β_1
Dry	-0.03	1.13
Moist	-0.79	1.02
Wet	-0.96	1.04
Snow/frost	-1.00	0.91
Ice	-0.97	0.97
Slush	-1.01	0.96

Look up the Teconer RCM411 road status observation and select the coefficients accordingly.

https://doi.org/10.1371/journal.pone.0211702.t006

To summarize, the inference models captured the effect of the environmental conditions on the mobile sensor observations. Dry, moist, wet, snowy, icy and slushy driving conditions all made subtle changes to the mobile observations, while a small but significant degradation was evident from a yearly trend. Further, a linear calibration, whose coefficients were estimated by fitting a linear mixed model, reduced the mobile sensor error by ca. 40%.

Conclusion

Mobile, vehicle-installed sensors and RWS networks can together provide denser and higher quality information of challenging driving conditions than either alone. First, this study analyzed the observations of mobile (Teconer RCM411, RTS411) and RWS (Vaisala DRS511, DST111, DSC111) road weather sensors and identified conditions and factors that affect how their measurements differ from each other. Second, this study presented a novel calibration method for the mobile sensors. A straightforward linear Eq (5) adjusted the mobile road surface temperature observations to be consistent with the stationary ones, halving the MSE between the mobile and the RWS observations.

The analysis and the calibration were based on rendezvous model sensor fusion, which considers spatio-temporally co-located observations of mobile and RWS sensors. Further, both analysis and calibration used linear mixed models to compare the mobile and RWS observations, taking into account the prevailing driving conditions as well as other environmental factors. The analysis revealed the following:

1. In dry conditions, at 0˚C surface temperature, with no water, ice, or snow deposits on the road, the mobile observations are on average 0.62 degrees higher than those of the RWS.
2. Different driving conditions, indicated by the RWS road status variable, have a significant effect on the difference. For example, mobile observations are a further 0.46˚C higher than RWS observations in wet conditions, totalling over 1˚C above RWS observations.
3. Model A4 indicates that the optical Vaisala DST111 RWS sensor gives on average 1.2˚C lower readings than the asphalt embedded Vaisala DRS511 in surface temperatures above 4˚C.
4. Mobile Teconer RTS411 sensor observations fall by ca. 0.2˚C year-by-year, justifying recurrent recalibrations.

Linear mixed models assume the general noise term is Gaussian. Model diagnostics revealed that while moderately well adhering to this assumption, the residuals of the inference models suggest the models could be further improved by replacing the Gaussian noise term distribution with one thicker tailed. For future work, we suggest embracing the Bayesian inference framework, with the noise term following e.g. Student's t distribution. Such an approach may reduce the relatively large standard error of the reported intercept estimates of the inference models. Further, the MSE's of the predictive models may be further lowered.

The sensor calibration method proposed in this study, based on sensor fusion and linear mixed effect models, is computationally very light. The linear calibration function can be embedded directly in the sensors. Further, the calibration method can be further utilized in other application areas such as air quality, temperature, and usage monitoring in smart building scenarios.

Supporting information

S1 Fig. Residual plot of the inference model.
(TIF)

S2 Fig. Q-Q plot of a sample of 100 residuals.
(TIF)

S3 Fig. Q-Q plot of the RWS sensor calibration level random effect.
(TIF)

S4 Fig. Q-Q plot of the mobile sensor calibration level random effect.
(TIF)

Author Contributions

Conceptualization: Lauri Lovén, Virve Karsisto, Heikki Järvinen, Mikko J. Sillanpää, Susanna Pirttikangas, Jukka Riekki.

Data curation: Lauri Lovén, Virve Karsisto.

Formal analysis: Lauri Lovén.

Funding acquisition: Mikko J. Sillanpää, Susanna Pirttikangas, Jukka Riekki.

Investigation: Lauri Lovén, Virve Karsisto.

Methodology: Lauri Lovén, Virve Karsisto, Mikko J. Sillanpää.

Project administration: Lauri Lovén, Virve Karsisto.

Resources: Lauri Lovén, Virve Karsisto.

Software: Lauri Lovén, Virve Karsisto.

Supervision: Heikki Järvinen, Mikko J. Sillanpää, Susanna Pirttikangas, Jukka Riekki.

Validation: Lauri Lovén, Virve Karsisto.

Visualization: Lauri Lovén, Virve Karsisto.

Writing – original draft: Lauri Lovén, Virve Karsisto.

Writing – review & editing: Lauri Lovén, Virve Karsisto, Heikki Järvinen, Mikko J. Sillanpää, Teemu Leppänen, Ella Peltonen, Susanna Pirttikangas, Jukka Riekki.

References

1. Haavasoja T, Nylander J, Nylander P. Experiences of mobile road condition monitoring. In: Proceedings of 16th International Road Weather Conference (SIRWEC). Helsinki, Finland; 2012. p. 23–25.
2. Pilli-Sihvola Y. Floating car road weather information monitoring system. Transportation Research Record: Journal of the Transportation Research Board. 2001; 1741:3–5. https://doi.org/10.3141/1741-01.
3. Stern AD, Shah VP, Biesecker KJ, Yeung C, Pisano PA, Pol JS. Vehicles as mobile sensing platforms for meteorological observations: a first look. In: Proceedings of the 87th AMS Annual Meeting, 23rd Conference on IIPS. San Antonio, US: American Meteorological Society; 2007. p. 1–10.
4. Koller D, Hersey BW, Mewes J, McCellan AK. New methods to quality check road and weather observations for vehicle data. Transportation Research Circular. 2012;April:349–362.
5. Malmivuo M. Comparison study of mobile optical friction and temperature meters 2013. Finnish Transport Agency; 2013.
6. Marjovi A, Arfire A, Martinoli A. Extending urban air quality maps beyond the coverage of a mobile sensor network: data sources, methods, and performance evaluation. Proceedings of the 2017 International Conference on Embedded Wireless Systems and Networks. 2017; p. 12–23.
7. Perttunen M, Mazhelis O, Cong F, Kauppila M, Leppänen T, Kantola J, et al. Distributed road surface condition monitoring. In: Hsu CH, Yang LT, Ma J, Zhu C, editors. Proceedings of the 8th International Conference on Ubiquitous Intelligence and Computing (UIC2011), LNCS, vol. 6905. Banff, Canada: Springer; 2011. p. 64–78.

8. Xiang Y, Bai L, Piedrahita R, Dick RP, Lv Q, Hannigan M, et al. Collaborative calibration and sensor placement for mobile sensor networks. In: Proceedings of the 11th international conference on Information Processing in Sensor Networks—IPSN'12. Beijing, China: IEEE; 2012. p. 73–83.
9. Hasenfratz D, Saukh O, Thiele L. On-the-fly calibration of low-cost gas sensors. In: Picco GP, Heinzelman W, editors. Proceedings of the 9th European Conference on Wireless Sensor Networks (EWSN2012), LNCS, volume 7158. Berlin Heidelberg: Springer; 2012. p. 228–244.
10. Saukh O, Hasenfratz D, Thiele L. Reducing multi-hop calibration errors in large-scale mobile sensor networks. In: Proceedings of the 14th International Conference on Information Processing in Sensor Networks—IPSN'15. Seattle, Washington: ACM; 2015. p. 274–285.
11. Arfire A, Marjovi A, Martinoli A. Model-based rendezvous calibration of mobile sensor networks for monitoring air quality. In: 2015 IEEE Sensors. Busan, South Korea: IEEE; 2015. p. 1–4.
12. Dorffer C, Puigt M, Delmaire G, Roussel G. Blind mobile sensor calibration using an informed nonnegative matrix factorization with a relaxed rendezvous model. In: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). vol. 2. Shanghai, China: IEEE; 2016. p. 2941–2945.
13. Li Q, Chen L, Li M, Shaw SL, Nüchter A. A sensor-fusion drivable-region and lane-detection system for autonomous vehicle navigation in challenging road scenarios. IEEE Transactions on Vehicular Technology. 2014; 63(2):540–555. https://doi.org/10.1109/TVT.2013.2281199
14. Ghose A, Biswas P, Bhaumik C, Sharma M, Pal A, Jha A. Road condition monitoring and alert application. In: 2012 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops). Lugano, Switzerland: IEEE; 2012. p. 489–491.
15. Vaisala. DRS511 Road Sensor, Product data sheet; 2014.
16. Vaisala. Remote Surface Temperature Sensor DST111, Product data sheet; 2017.
17. Vaisala. Remote Road Surface State Sensor DSC111, Product data sheet; 2010.
18. Bridge P. Non-invasive road weather sensors. In: Fourth National Conference on Surface Transportation Weather, Seventh International Symposium on Snow Removal and Ice Control Technology. June. Indianapolis, Indiana, US: Transportation Research Board of the National Academies; 2008. p. 407–415.
19. Robinson GK. That BLUP is a good thing: The estimation of random effects. Statistical Science. 1991; 6(1):15–51. https://doi.org/10.1214/ss/1177011933
20. West BT, Welch KB, Galecki AT. Linear mixed models: A practical guide using statistical software. Boca Raton, FL, USA: Chapman & Hall/CRC; 2007.
21. Greven S, Kneib T. On the behaviour of marginal and conditional AIC in linear mixed models. Biometrika. 2010; 97(4):773–789. https://doi.org/10.1093/biomet/asq042
22. Marquardt DW. Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. Technometrics. 1970; 12:591–612. https://doi.org/10.1080/00401706.1970.10488699

Reprinted, with permission, from
Weather and forecasting, 34, no. 3, 539–558,
doi:10.1175/WAF-D-18-0167.1

IV

Verification of Road Surface Temperature Forecasts Assimilating Data from Mobile Sensors

VIRVE KARSISTO

Finnish Meteorological Institute, and University of Helsinki, Helsinki, Finland

LAURI LOVÉN

University of Oulu, Oulu, Finland

(Manuscript received 28 September 2018, in final form 24 January 2019)

ABSTRACT

The advances in communication technologies have made it possible to gather road condition information from moving vehicles in real time. However, data quality must be assessed and its effects on the road weather forecasts analyzed before using the new data as input in forecasting systems. Road surface temperature forecasts assimilating mobile observations in the initialization were verified in this study. In addition to using measured values directly, different statistical corrections were applied to the mobile observations before using them in the road weather model. The verification results are compared to a control run without surface temperature measurements and to a control run that utilized interpolated values from surrounding road weather stations. Simulations were done for the period 12 October 2017–30 April 2018 for stationary road weather station points in southern Finland. Road surface temperature observations from the stations were used in the forecast verification. According to the results, the mobile observations improved the accuracy of road surface temperature forecasts when compared to the first control run. The statistical correction methods had a positive effect on forecast accuracy during the winter, but the effect varied during spring when the daily temperature variation was strong. In the winter season, the forecasts based on the interpolated road surface temperature values and the forecasts utilizing mobile observations with statistical correction had comparable accuracy. However, the tested area has high road weather station density and not much elevation variation, so results might have been different in more varying terrain.

1. Introduction

Vehicles providing road condition information will have an important role in the future intelligent transport systems (ITS). They enable detailed knowledge of road conditions along the road network, which is critical to ensure the safety of autonomous vehicles in winter conditions. Traditionally, road condition forecasts have been based on the observations of stationary road weather stations (RWS), but these are sparsely located especially in less populated areas. Road surface temperature is very dependent on local conditions, especially on the openness of the location. There can be even 10°C surface temperature variability at different parts of the road network (Shao et al. 1996; Bogren et al. 2000). Local temperature anomalies may remain unnoticed if the monitoring network is too sparse. However, installing traditional RWSs to also cover the smaller roads would be very expensive. Vehicle-based observations enable much more detailed road condition monitoring in a cost-effective way. In an ideal situation, the observations could be obtained in a great volume from common vehicles, but a lot of data can be also gathered using selected vehicles with a lot of driving kilometers such as buses and taxis.

A prerequisite for accurate forecasts is information about the current conditions at the forecast area. It is important that the initial conditions in the forecasting model represent real conditions, because even small deviations can change the forecast. Mobile observations have the potential to improve the forecast accuracy for road stretches without RWSs. Accurate road condition forecasts enable well-timed road maintenance actions, which reduce the road accident risk and potentially lead to economical savings as unnecessary actions can be avoided. For example, roads can be salted beforehand if the forecast correctly estimates when the road surface temperature will drop below the freezing point. This prevents

Corresponding author: Virve Karsisto, virve.karsisto@fmi.fi

DOI: 10.1175/WAF-D-18-0167.1

the roads from freezing and becoming dangerous for drivers. However, the roads should not be salted too early, as traffic causes the salt to wear off from the road. More accurate forecasts also produce less false alarms, and there will be fewer cases when the roads are salted although the surface temperature remains above 0°C. Accurate forecasts also enable detailed warnings about dangerous road conditions beforehand so that road users can select their route and adjust their driving behavior accordingly.

Thermal mapping has been commonly used to estimate surface temperatures on road sections without RWS measurements (Thornes 1991; Shao et al. 1997; Gustavsson 1999). The method is based on the idea that the spatial behavior of surface temperature is reproduced in similar weather situations. The surface temperature along the target road network is measured several times using a measurement instrument attached to a vehicle. The measured temperature variability can be used later to determine the surface temperature along the road network. However, the increasing availability of mobile observations will make it possible to obtain real-time observations along roads in the future. Modern cars are equipped with a multitude of sensors and can provide road condition and weather information. Several earlier studies discuss the potential use of cars as sensors (Allegretti and Bertoldo 2014; Petty and Mahoney 2007). A disadvantage of thermal mapping and mobile observations is that they provide only a snapshot of the current situation and do not tell about the temporal behavior of surface temperature (Chapman and Thornes 2006). This might be changed if the majority of cars would provide observations in the future, but currently the availability of data is still a problem. Obtaining data from the vehicles' internal communication systems is difficult, because in normal cars an unauthorized accessing of the car's data can even cause the rejection of the vehicle's warranty. Data can be more easily collected from heavy vehicles, but there still might not be detailed instructions for how to interpret the data contents.

The aim of the research presented in this paper was to find out if the road surface temperature forecast accuracy can be improved by using observations from vehicles. A disadvantage of mobile observations is that they are more exposed to disturbances, such as heating from the vehicle or drifting snow. The quality of the measurements made with mobile sensors might not be as high as measurements made at RWSs. The data quality of mobile observations should be studied before using them for verification, postprocessing, or assimilation in a forecasting system. It should also be ensured that the forecast accuracy increases when the forecast utilizes the new data.

This study presents verification results of road surface temperature forecasts that assimilate observations made with an optical Teconer RTS411 instrument (Haavasoja et al. 2012). The next section describes the Finnish Meteorological Institute's (FMI) road weather model. Section 3 gives information of the datasets used in the simulations and explains the statistical calibration methods applied to the mobile observations. Results of the forecast verification are described in section 4. Section 5 contains a discussion, conclusions, and suggestions for further research.

2. Road weather model

The Finnish Meteorological Institute's road weather model (RWM) has been in operation since 2000 (Kangas et al. 2015). The model is a one-dimensional heat balance model and aims to predict the temperature at the surface. Other model outputs are the amounts of water, ice, snow, and frost on the road, friction, road condition, and traffic index (Fig. 1). The road condition describes the status of the road, which can be for example "wet" or "icy." The traffic index depicts the overall driving conditions.

The calculation of surface temperature is based on the surface energy balance (Brutsaert 1984):

$$G = I_{\text{net}} - H - \text{LE} + \text{Tr}, \tag{1}$$

where G is heat flux into the ground, I_{net} is net radiation at the surface, H is sensible heat flux, LE is latent heat flux, and Tr describes heating caused by traffic. Figure 2 illustrates the model energy fluxes. The ground is divided into several layers, which are also shown in Fig. 2. The calculation of heat transfer between the ground layers is based on (Patankar 1980):

$$\rho_g c_g \frac{\delta T(z,t)}{\delta t} = \frac{\delta}{\delta z} K \frac{\delta T(z,t)}{\delta z}, \tag{2}$$

where T is temperature, z is vertical distance in the ground, t is time, K is heat conductivity, ρ_g is density, and c_g is specific heat capacity of the ground. To calculate the temperature at the next time step, Eq. (2) is integrated over the time step and the volume of the layer and solved using the forward difference explicit method:

$$T_i^{j+1} = T_i + \frac{1}{\rho_g c_g \dfrac{z_{i+1} - z_{i-1}}{2\Delta t}} \times \left(K_i \frac{T_{i+1}^j - T_i^j}{z_{i+1} - z_i} - K_{i-1} \frac{T_i^j - T_{i-1}^j}{z_i - z_{i-1}} \right), \tag{3}$$

where index i refers to the ground layer and index j to time, Δt is the model time step, and K_i means heat conductivity between layers $i + 1$ and i.

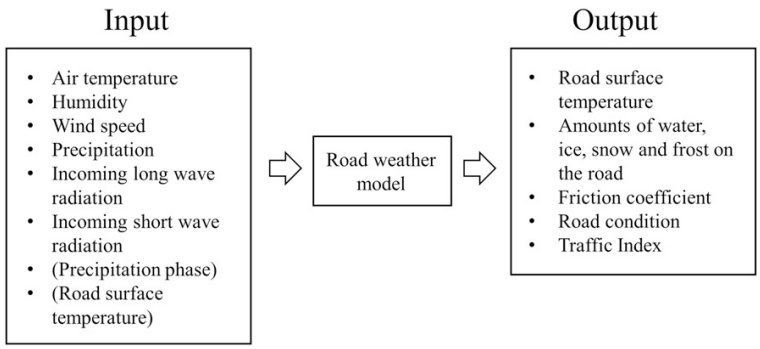

FIG. 1. Road weather model input and output values. Precipitation phase and road surface temperature are optional. Input road surface temperature is used only in model initialization.

More detailed descriptions of the model physics are given by Kangas et al. (2015) and Karsisto et al. (2017). The model has been recently updated, and the model version used has some differences when compared to the models described in the publications. One of the most influential changes is the model time-step reduction to 30 s for the whole simulation. The original model had a time step of 5 min. Another major change was done to the ground temperature profile calculation. In the original model, the temperature profile for the next time step was determined iteratively by solving the temperature for each layer simultaneously (Kangas et al. 2015). In the new model version, new temperature for each layer is calculated directly implementing Eq. (3) similarly as in the METRo model (Crevier and Delage 2001).

The RWM requires atmospheric values (air temperature, humidity, wind speed, precipitation, and shortwave and longwave radiation) as input. The model run consists of the initialization phase and the forecast phase. In the initialization phase, the atmospheric values are obtained from the observations, and the model is run for circa 2-day period to get a good initial state for the ground temperature profile. However, the radiation parameters are obtained from the 3D numerical weather prediction model (NWP) due to the lack of observations. In the forecast phase, all the aforementioned atmospheric values are obtained from the NWP forecast. The FMI RWM is not a stand-alone model, but it relies on the NWP model in regard to the large-scale weather prediction. Using the values from the NWP forecast, the model aims to predict the road conditions at the forecast point.

The RWM uses mobile road surface temperature observations in the initialization by the coupling method (Crevier and Delage 2001; Karsisto et al. 2016). The method iteratively determines a correction coefficient for the radiation so that the surface temperature in the simulation fits to the observed road surface temperature. This iterative radiation adjustment period was set to 3 h in the model. The coefficient is given for either shortwave or longwave radiation depending on which has the higher value at the start of the coupling period. The effect of the coefficient decreases exponentially as the forecast advances. The coupling method makes it possible to use even a single road surface temperature observation efficiently.

3. Data and methods

The road surface temperature forecasts were produced for the RWS points. The forecasts were not run in

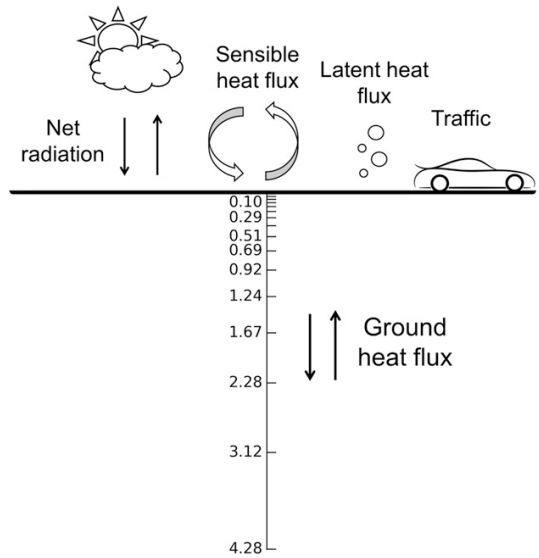

FIG. 2. The depths of the ground layers in the road weather model (m) and a schematic representation of energy fluxes.

real time but as hindcasts afterward. Surface temperature observations from the stations were used for forecast verification. The results are compared to two types of control runs that did not use the mobile observations to find out the effect of the new data on the forecast. The first type of control run did not use any kind of road surface temperature observations. The second type utilized interpolated road surface temperature values, which were obtained by using observations from independent RWSs, as in a leave-one-out cross-validation experiment. The first control run simulates a forecast situation where the road weather station network is very sparse and the second run a situation with multiple stations nearby the forecast target point.

a. Mobile measurements

Teconer RCM411 and RTS411 are optical measurement instruments that can be attached to vehicles (Haavasoja et al. 2012). RCM411 measures optically water layer thickness, road condition, and friction. The functionality of the device is based on the absorption of radiation at near-infrared wavelengths. Dry, wet, and icy roads reflect the radiation differently at these wavelengths, and thus the road condition can be determined. RTS411 measures road surface temperature by detecting infrared radiation. The measurements are done every second. The mobile measurements used in the simulations were from Teconer devices attached to vehicles that passed selected RWS points in southern Finland. An average of road surface temperature measurements within 50 m from the RWS during a pass was used as the RWM model initialization value.

Before this study, statistical analysis was performed for RWS and RTS411 road surface temperature measurements to find out possible differences. Mobile sensors are more exposed to disturbances than RWS, like heating from the car and drifting snow. For example, the infrared radiation emitted by the vehicle can reflect from the road to the optical surface temperature sensor and thus cause a warm bias. The statistical analysis was based on linear mixed effect models and used data from three winter periods between years 2014 and 2017. The data were gathered with a rendezvous sensor fusion method similarly as in this study, identifying rendezvous cases where a vehicle passed an RWS. The averages of road surface temperature measurements within a 50-m radius of the RWSs during each rendezvous were calculated. The analysis is covered in detail by Lovén et al. (2019). According to the results, the Teconer observations are on average 0.62°C warmer than the RWS observations in dry conditions and at 0°C temperature. The temperature difference was found to be dependent also on the road status (dry, wet, or icy) and the individual device. Water, ice, and snow on the road surface can cause the optical sensor to measure the temperature of the substance rather than the actual road surface temperature. This causes the difference in comparison with the RWS measurements, because at most RWSs in Finland the surface temperature observations are done with asphalt embedded sensors. The device dependency can be partially caused by the device's location in the vehicle. For example, nearby warm engine bodies can emit infrared radiation that disturbs the surface temperature measurements when this radiation is reflected from the road surface to the sensor.

b. Statistical calibration methods

Three sensor calibration methods for the mobile surface temperature data were tested, which were all based on linear statistical models. The first one used the calibration provided by a linear mixed effect model (Robinson 1991; West et al. 2007) constructed by Lovén et al. (2019). The model was fitted with Teconer and RWS observations from winter periods between the years 2014 and 2017. Winter period 2017–18 used in this study was not included in the model development. The model assumed that RWS and mobile road surface temperature observations are linearly related, with the road status (i.e., dry, wet, icy, etc.) affecting the coefficients of the linear relationship. Further, the model assumed that the current calibration levels of both RWS and mobile sensors (i.e., their reading when actual road surface temperature was 0°C), each follow individual Gaussian distributions:

$$\mathbf{T}_{s,\mathrm{RWS}} = \mathbf{X}_{\mathrm{Mob}}\boldsymbol{\beta} + \mathbf{Z}_{\mathrm{id}}\mathbf{b} + \boldsymbol{\varepsilon}. \tag{4}$$

Here, $\mathbf{T}_{s,\mathrm{RWS}}$ corresponds to the RWS road surface temperatures, $\mathbf{X}_{\mathrm{Mob}}$ is a design matrix of the mobile road status and surface temperature observations as well as the overall intercept, corresponding to dry road status, while $\boldsymbol{\beta}$ is the fixed effect coefficient. The term $\mathbf{Z}_{\mathrm{id}}$ includes dummy sensor identifiers for both the mobile and the RWS sensors, with the random coefficient $\mathbf{b} = (b_1, b_2)$, corresponding to mobile (b_1) and RWS (b_2) calibration levels, distributed normally as $b_1 \sim N(0, I\sigma_1^2)$ and $b_1 \sim N(0, I\sigma_2^2)$. The noise term $\boldsymbol{\varepsilon}$ is also assumed Gaussian with zero mean.

The calibration in Eq. (5), corresponding to the expectation of $\mathbf{T}_{s,\mathrm{RWS}}$, adjusts the mobile observations with a linear function of the fixed coefficients:

$$T_{s,\mathrm{Adj}} = \beta_0 + \beta_1 T_{s,\mathrm{Mob}}. \tag{5}$$

Here, $T_{s,\mathrm{Adj}}$ is the calibrated mobile road surface temperature, while $T_{s,\mathrm{Mob}}$ is the original mobile observation,

TABLE 1. Status-based coefficients used in Eqs. (2)–(4).

Road status	β_0	β_1	β_2	β_4	β_5
Dry	0.03	1.13	1.08	0.00	1.13
Moist	−0.79	1.02	1.08	−0.72	1.02
Wet	−0.96	1.04	1.08	−0.90	1.04
Snow/frost	−1.00	0.91	1.08	−0.86	0.86
Ice	−0.97	0.97	1.08	−1.07	0.95
Slush	−1.01	0.96	1.08	−1.02	0.98

TABLE 2. Device-based coefficients used in Eqs. (3) and (4).

Device	β_3	β_6
A	−0.60	−0.25
B	−1.08	−0.43
C	0.21	0.69
D	−0.42	0.12
E	−0.87	−0.20
F	−1.67	−1.11
G	−1.43	−1.02

β_0 is the estimated road status intercepts, and β_1 is the estimated interaction coefficient between the road status and the mobile surface temperature. The second and third models modified Eq. (3) by including individual intercepts for the mobile sensors instead of assuming them to be normally distributed. Further, the second model left out the coefficients for the road status observations. Corresponding calibration equations are provided below as Eqs. (6) and (7), while the estimated fixed-effect β coefficients for all calibration equations, obtained by fitting the models, can be found in Tables 1 and 2:

$$T_{s,\mathrm{Adj}} = \beta_2 T_{s,\mathrm{Mob}} + \beta_3, \tag{6}$$

$$T_{s,\mathrm{Adj}} = \beta_4 + \beta_5 T_{s,\mathrm{Mob}} + \beta_6. \tag{7}$$

In the following sections, the first model [Eq. (5)] is referred as "status-based correction," the second model [Eq. (6)] as "device-based correction," and the third model [Eq. (7)] as "status- and device-based correction."

c. *Kriging and forecast data*

The RWM used interpolated air temperature, humidity, and wind speed values as atmospheric forcing in the initialization phase. The values were interpolated from surface synoptic observations (SYNOP) weather stations to a 1-km grid with a universal kriging method (Cressie 1993; Aalto et al. 2013). Elevation data, lake percentage, and sea percentage were used as explanatory variables in the interpolation. Lake percentage means the percentage of lake surface in the grid cell, and correspondingly, sea percentage means the percentage of sea surface in the grid cell. Altitude and water bodies cause spatial trends in the air temperature, humidity, and wind speed values. For example, air temperature decreases with altitude and the large heat capacity of water may cause it to change slower near lakes. These trends are removed from the data in the kriging analysis, and the spatial autocorrelation is calculated from the residuals. The gridded values were interpolated from the grid to the simulation points with bilinear interpolation. Hourly precipitation values were obtained from weather radar observations. Interpolated road surface temperature values for the second control run were also obtained with the kriging method by using observations from road weather stations. The interpolation was done to a 10-km grid and was performed separately for each simulation point so that the road surface temperature observations from that point were not included in the calculation. After the kriging analysis, the value for the simulation point was obtained by bilinear interpolation from the nearest grid points.

The atmospheric values used in the forecast phase of the model were obtained from the forecasts made with the HARMONIE-AROME (HIRLAM–ALADIN Research on Mesoscale Operational NWP in Euromed–Applications of Research to Operations at Mesoscale) model configuration. The model is a convection-permitting nonhydrostatic NWP (Bengtsson et al. 2017). The forecasts were run four times a day for the Scandinavia region with 2.5-km resolution by FMI. The analysis times of the runs were 0000, 0600, 1200, and 1800 UTC. Local time in Finland is UTC +2 h during wintertime and UTC +3 h during summertime. The simulations were done for road weather station points located on the Helsinki–Turku motorway or on the nearby roads (Fig. 3, Table 3). Road surface temperature measurements used in the forecast verification were done with asphalt-embedded Vaisala DRS511 (Vaisala 2001). As the measurement interval of RWSs is 5–10 min, the nearest measurement of the verification time was used. The used NWP data, road weather station measurements, and hourly radar values are available in a public repository (Karsisto 2018).

d. *Model run construction*

The model runs consist of a 52-h initialization phase and a 21-h forecast phase (Fig. 4). In the initialization phase, the air temperature, wind speed, and humidity values were obtained from the kriging data. Radar measurements were used to obtain hourly precipitation values, but the precipitation phase was taken from HARMONIE forecasts so that the values from the first six forecast hours from each run formed a continuous time series. Shortwave and longwave radiation for the initialization were obtained similarly from forecasts. The forecast phase starts three hours after the

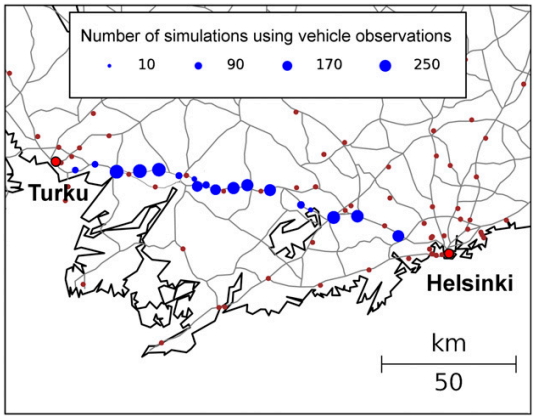

FIG. 3. Road weather station points in Southern Finland. Blue circles represent points included in the forecast verification in this study. The size of the circle shows how many simulations in that point were able to utilize mobile observations. Brown dots show other road weather stations. All of the stations along the Helsinki–Turku motorway did not have observations during the studied period.

TABLE 3. List of road weather stations included in the study.

Latitude (°N)	Longitude (°E)	Name
60.42402	22.40233	Kaarina
60.44583	22.53482	Piikkiö
60.4237	22.68302	Paimio
60.42836	22.84023	Pitkäporras
60.43515	22.96914	Hajala
60.41729	23.10632	Halikko
60.4074	23.21123	Haukkala
60.38308	23.23238	Tupuri
60.38867	23.29192	Muurla
60.37346	23.35686	Lakiamäki
60.38049	23.4779	Kruusila
60.39117	23.5713	Suomusjärvi
60.37624	23.72575	Lahnajärvi
60.32855	23.93642	Pitkämäki
60.31271	24.00195	Karnainen
60.28811	24.15869	Lehmijärvi
60.29328	24.31916	Palojärvi
60.22805	24.59641	Nupuri

HARMONIE analysis time and utilizes the latest HARMONIE forecast. If there is a mobile surface temperature observation available within this 3-h period, it is utilized in the simulation with the coupling method. The simulation used the latest observation if there were multiple observations available at the forecast point during this time window. The second control run used interpolated road surface temperature values for the corresponding times so that it would not have an advantage by using more recently made observations. The coupling method's radiation adjustment period was set to 3 h. Simulations were done four times a day for the period 12 October 2017–30 April 2018. As forecasts start three hours after the HARMONIE analysis time, the corresponding forecast starting times were 0300, 0900, 1500, and 2100 UTC. The simulations were not done as real-time forecasts but as hindcasts afterward. The output surface temperature values for the forecast phase are available also in the public repository (Karsisto 2018).

e. *Verification methods*

Verification results were calculated separately for the periods 12 October 2017–17 February 2018 and 18 February–30 April 2018 because the daily surface temperature variation was much stronger during the springtime than winter. This caused great differences in the verification results for the two periods. The date of 18 February was chosen for the date to separate the two periods because after that the average daily surface temperature variation over the simulation points used was almost consistently above 10°C. Only the cases where there was a mobile observation available within three hours before the start of the forecast phase were included in the verification. The verification concentrates on the first 10 forecast hours because the effect of the coupling method decreases as the forecast advances.

The bootstrap method was used to calculate whether the differences in the verification scores between runs were statistically significant (Efron and Tibshirani 1993; Hogan and Mason 2011). Multiple samples from the dataset were generated, and the differences between the scores were calculated from each sample. The 95% confidence interval was calculated from the distribution of differences. If 0°C was not in the obtained range, the differences between models were considered statistically significant with 95% confidence (Hogan and Mason 2011). The upper and lower limits of the range were rounded to two decimals before the check except in the case of the correlation coefficient where four decimals were used due to small differences. In this study the sample size used was the same

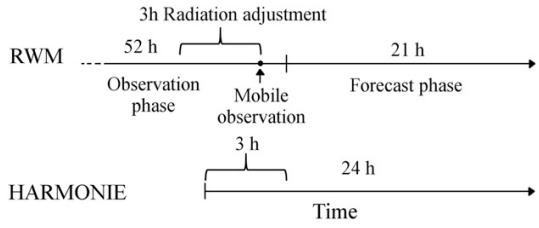

FIG. 4. The constructions of a model run.

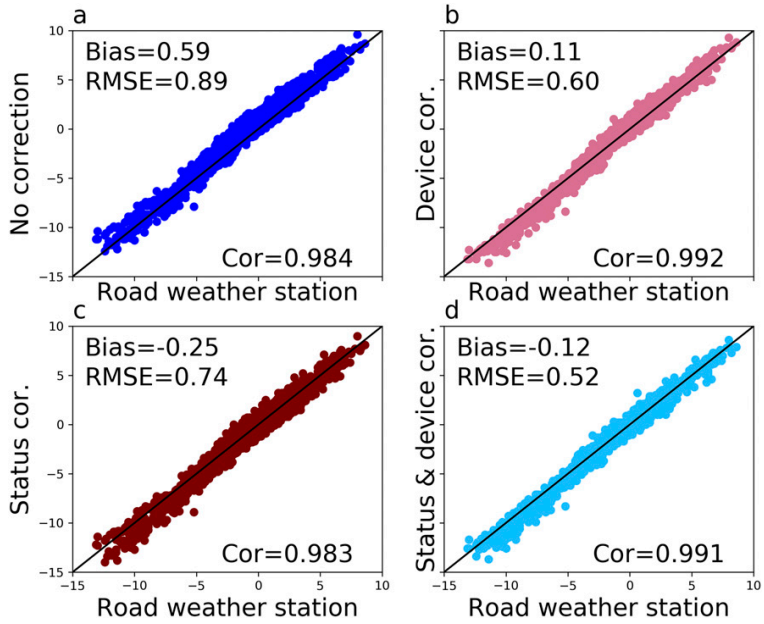

FIG. 5. Scatterplots of road surface temperature measured at road weather stations (x axis) compared to mobile observations used in the initialization of model runs during the midwinter period (y axis). Bias, RMSE, and correlation coefficient values are shown in the corners of each panel. (a) Mobile observations without statistical correction, (b) device-based correction, (c) status-based correction, and (d) status- and device-dependent correction methods.

as the size of the dataset used, and the number of generated samples was 10 000. Samples were generated with replacement.

4. Results

a. Input mobile measurements

The mobile road surface temperature measurements used in the RWM initialization were compared to the road weather station measurements to find out the initial error in the model runs. In Fig. 5, there are four scatterplots with surface temperature measured at RWSs on the x axis and mobile measurements on the y axis. Different adjustment equations are applied to the mobile measurements in Figs. 5b–d. According to the figure, the Teconer observations have a warm bias compared to the RWS observations during the winter period. The warm bias is reduced when a statistical correction is applied to the Teconer measurements. The bias even turns negative when the correction based on road status is applied. In addition, all of the correction methods seem to cause too cold values in cases where RWS surface temperature was below −5°C. The correction methods are more in line with RWS observations in temperatures over 0°C, although there is some warm bias when the device-based correction is applied. The figure shows also that all correction methods reduce the root-mean-square error (RMSE) of Teconer observations. The best results are obtained when the status- and device-based correction method is used (Fig. 5d). The RMSE and bias differences between correction methods were determined to be statistically significant with 95% confidence by using the bootstrap method as explained in section 3e. Differences between correlation coefficients were statistically significant in all cases except between the status-based correction method and in the case where no correction was used.

Figure 6 contains similar scatterplots as Fig. 5 but for the spring period. It shows that during spring time the Teconer observations are also usually warmer than RWS observations in temperatures below −5°C. The correction methods again reduce the warm bias of Teconer observations in these temperatures. Conversely, RWSs usually give warmer surface temperature measurements than Teconer instruments when the RWS observation is over 5°C. This is probably caused by the asphalt-embedded sensor heating too much during sunny days. The correction methods taking into account the road status somewhat reduce the warm bias, although the Teconer instrument might actually be more correct

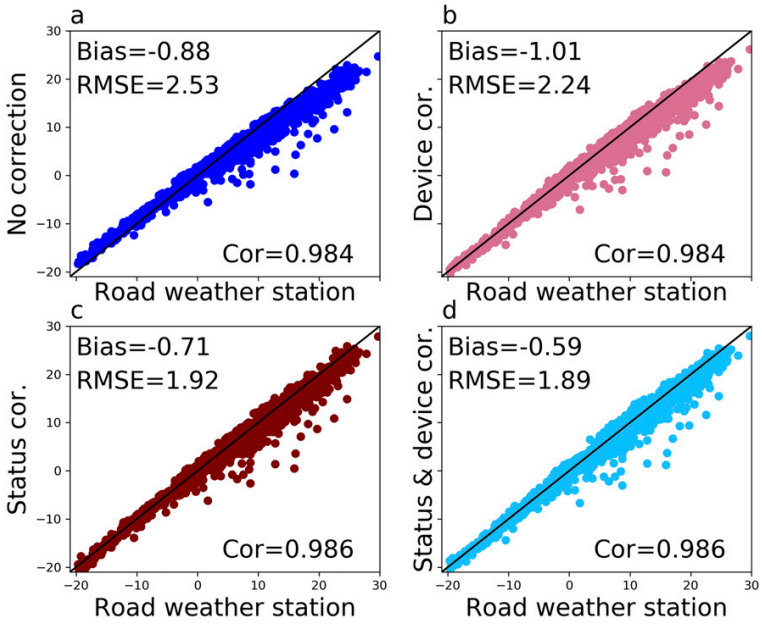

FIG. 6. As in Fig. 5, but for the spring period.

in the sunny cases by showing colder temperatures. The asphalt sensor overheating is probably the reason for the higher RMSE values during the spring period than during the winter period. To test this hypothesis, the Teconer observations were compared to the road surface temperature measurements done at stations equipped with the Vaisala DST111. DST111 measures surface temperature optically similar to the Teconer RTS411, so it should not have the overheating issue. The measurements were done during time period from 1 October 2017 to 30 April 2018. For reference, the Teconer observations done with the same individual devices were compared to the measurements done with asphalt-embedded sensors at the RWSs. However, these measurements were from different stations than where the DST111 measurements were done. The Teconer observations clearly had more negative bias values when compared to measurements done with the asphalt-embedded sensors than when compared to the measurements done with DST111. Only cases where the RWS observation was above 5°C were included in the comparison. This result indicates that the reason for Teconer observations being colder than RWS measurements at high temperature values (Fig. 6) is the asphalt sensor overheating and is not an issue related to the Teconer devices.

Overall, the status- and device-based correction method (Fig. 6d) gave the best results also during the spring period. The statistical significance of the score differences between correction methods was calculated similarly as for the winter period. The RMSE and bias differences were determined to be statistically significant at 95% confidence in all cases expect between the RMSE values of the two methods utilizing status-based correction. Differences between correlation coefficients were statistically significant in all cases except when the device-based correction method was compared to the cases with no correction and between the correction methods utilizing the status-based correction.

b. *Interpolated values*

To better explain the forecast verification results, scatterplots in Fig. 7 compare the interpolated road surface temperature values used in the second control run to road weather station measurements. According to the figure, there is some cold bias in both the winter and spring periods. During the spring period, the RMSE value is considerably smaller than the values for statistically corrected mobile observations. Road surface temperature is measured with asphalt-embedded sensors at most road weather stations in Finland, so overheating in sunny weather affects the interpolated values. Consequently, the interpolated values are closer to RWS measurements than mobile measurements that are not affected by the overheating effect.

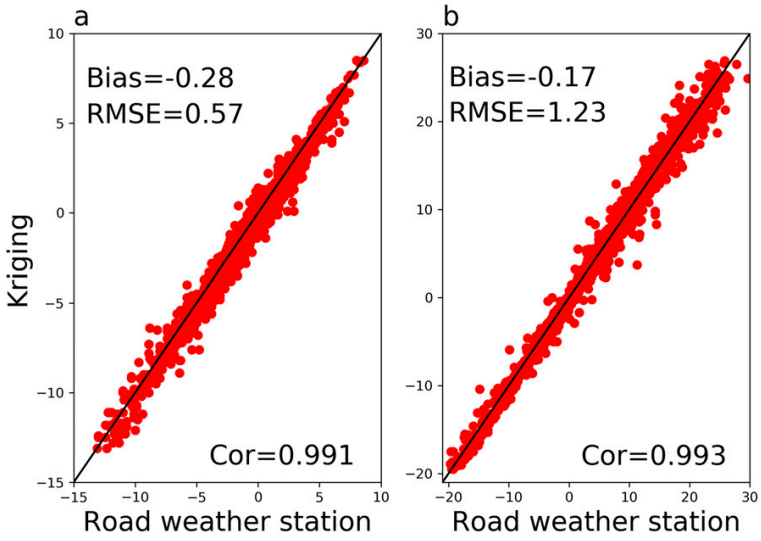

FIG. 7. Scatterplots of road surface temperature measured at road weather stations (x axis) compared to interpolated values from other road weather stations used in the initialization of the second control run (y axis) during the (a) midwinter period and (b) spring period. Bias, RMSE, and correlation coefficient values are shown in the corners of each panel.

c. *Case study*

One of the forecast cases is described in detail to demonstrate how the mobile observations contribute to the forecast. Figure 8 shows the modeled surface temperatures from all the forecast runs started at 0300 UTC 24 December for the Paimio station in addition to the RWS observations. The stars show the value used by the coupling method in the corresponding model run. All the simulations have the same modeled surface temperatures until the start of the coupling phase. In the coupling phase, the radiation correction coefficient is determined so that the modeled value fits to the observed value. The first control run that did not use the surface temperature observations and thus did not use the coupling method was too warm before the start of the forecast. The mobile observation is also too warm but brings the modeled surface temperature a little closer to the RWS value. Modifying the mobile observation with one of the three statistical corrections further decreases the modeled temperature. However, the kriged value in this case is too cold and causes also the first two forecast hours of the second control run to have too low temperature values. Nevertheless, as the observed temperature decreases, the run assimilating the kriged value is closest to the RWS observations at the third forecast hour. At the first two forecast hours, the runs using the mobile observation with a statistical correction are closest to the RWS observations. The RWS surface temperature starts increasing at 0600 UTC, and after two hours all the runs give too low surface temperatures because the temperature increases too slowly in the model. The differences between runs decrease gradually as the lead time increases.

d. *Forecast verification for all stations*

Only the cases where the RWS surface temperature was below 10°C were included in the forecast verification as we are mainly interested in near-zero temperatures. This also reduces the error caused by asphalt sensor overheating, although it probably still has an effect on the verification results. Figures 9–11 show RMSE and bias values of the forecasts for both winter and spring periods. The results for the spring period are separated by the forecast start time due to the strong daily surface temperature variations.

Using the mobile observations in the RWM initialization reduced the RMSE values at the beginning of the forecast for both verification periods, when compared to the first control run (Figs. 7a and 9). Using the mobile observations without a statistical correction causes a warm bias in the model during the winter period (Fig. 7b). This is in line with the fact that the initial surface temperature values measured with the Teconer instrument tend to be warmer than the RWS

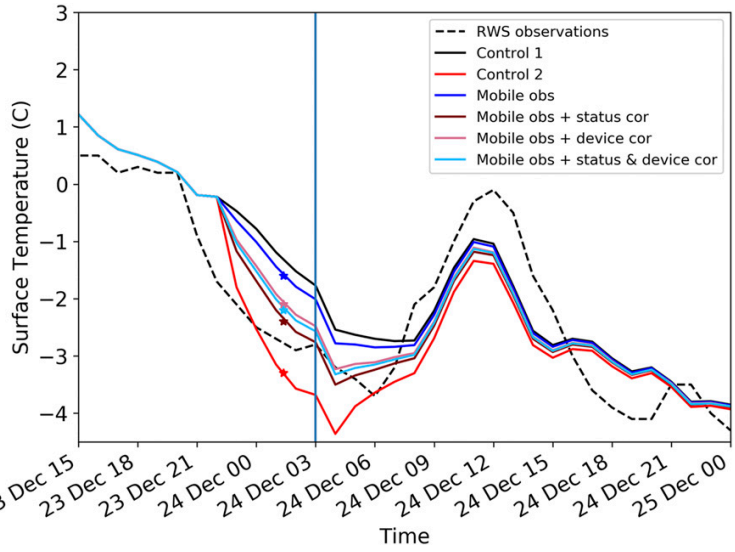

FIG. 8. Example of road surface temperature forecasts done for Paimio station for 24 December 2017. Observations are shown by the black dashed line. The black continuous line shows the control run without assimilated road surface temperatures, the red line is the control run with kriged and interpolated RWS temperatures, the blue line is the run using mobile observations, the brown line is the run using mobile observations with status-based correction, the pink line is the run using mobile observations with device-based correction, and the cyan line is the run using mobile observations with both status- and device-based correction. The vertical line shows where the initialization phase ends and the forecast phase starts. Stars show the surface temperature values assimilated in model runs by using the coupling method. The stars' colors correspond to the line colors of the model runs.

values during the winter (Fig. 5). Adding statistical correction to the Teconer observations reduced the warm bias. The negative bias of the correction method applying road status seen in the initial temperature values also causes negative bias in the beginning of the forecast during the winter period. Looking for both the bias and RMSE values for the winter period, the correction method applying both the status and device correction gives the best results of the runs utilizing mobile observations. However, the differences in the RMSE values were statistically significant with 95% confidence between these three runs only at the first forecast hour. Starting from the second forecast hour, the difference between the two runs utilizing the device-based correction was not found to be statistically significant with the bootstrap method. Starting from the fourth forecast hour, the differences between each of the three runs utilizing the statistical correction methods were not statistically significant with 95% confidence. (The results of the statistical significance tests can be found in Tables A2 and A3 in the appendix.)

The second control run utilizing the interpolated observations gives considerably smaller RMSE values than the first control run during the winter period (Fig. 7a). However, the negative bias seen in the initial values is present also during the first forecast hours (Fig. 7b). The RMSE values of the second control run are very close to the values of the runs using mobile observations with the device-based correction methods during winter. At the first two forecast hours the results are not different enough to reach the 95% significance level. The bias values of the second control run are very similar to the run using mobile observations with the status-based correction method.

The observation error (Fig. 6) and the forecast error for the model runs assimilating mobile observations with a statistical correction (Fig. 9) are about the same order of magnitude for wintertime. The same is true for the RMSE of the kriged surface temperature values (Fig. 7) and the RMSE for the runs assimilating them (Fig. 9). The forecast error is comparable to the observation uncertainty, and thus it is possible that the observation error has a dominant effect on the forecast accuracy. This was

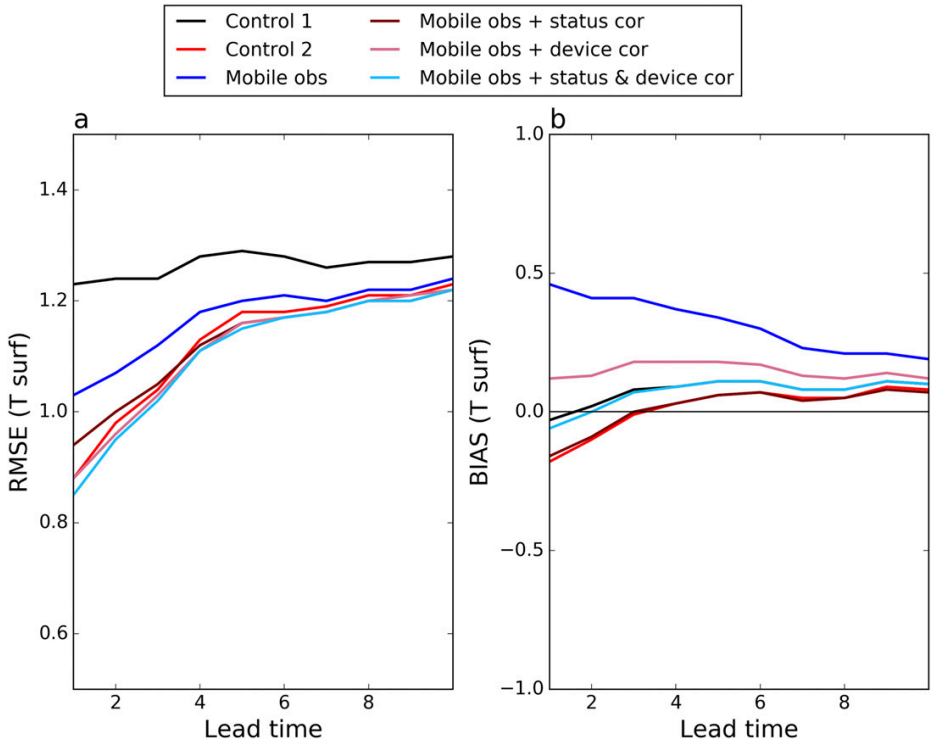

FIG. 9. (a) RMSE and (b) bias values of road surface temperature forecasts as function of forecast lead time. The results are calculated for the midwinter period and for all forecast starting times. Only the cases where the RWS surface temperature was below 10°C were included in the forecast verification. The black line shows results for the control run without assimilated road surface temperatures, the red line for the control run with kriged and interpolated RWS temperatures, the blue line for the run using mobile observations, the brown line for the run using mobile observations with status-based correction, the pink line for the run using mobile observations with device-based correction, and the cyan line for the run using mobile observations with both status- and device-based correction.

investigated by studying the results of a third control run that is not otherwise analyzed in this study. The model settings were similar than in the other runs, but the assimilated surface temperatures were taken from the RWS. Thus, the third control run did not involve error caused by surface temperature observation uncertainty, as the RWS observations are considered "truth" in this study. The third control run's RMSE values for the first two forecasts hours for the winter period were only slightly lower than for the model run assimilating mobile observations with the status- and device-based correction. Starting from the third forecast hour, the RMSE values of these runs were rather similar. It can be concluded that the forecast error is the same order of magnitude regardless of the observation uncertainty involved in statistically corrected mobile observations. This is the case also with kriged values. Uncertainty related to the forecasting model and other input values seems to also have a great effect on the forecast error. The effects of errors in the initial values on the forecast accuracy are complicated and do not directly accumulate with forecast error. The erroneous values can even improve the forecast in some cases if they compensate for some other error in the forecast.

During the springtime, the forecast accuracy was not as good as during the winter period according to the verification results (Figs. 10 and 11). The daily temperature variation was much stronger, and the RWM had difficulties in the forecasts of the daytime maximum and the nighttime minimum temperatures. The RMSE has a peak value at 1000 UTC in the forecasts started at 0300 UTC because the model greatly underestimated the daytime maximum surface temperatures (Fig. 10a). A part of the error might be caused by the asphalt sensor overheating too much during sunny days. In the forecasts started at 0900 UTC the RMSE is largest at the first

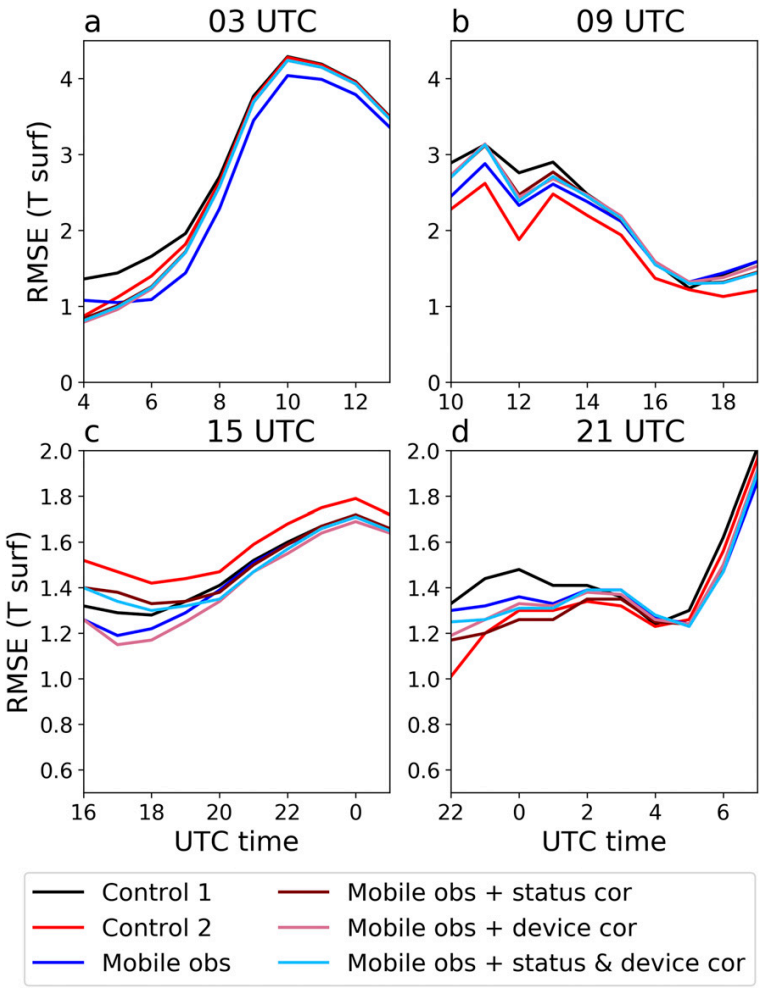

FIG. 10. RMSE values of road surface temperature forecasts as function of UTC time. The values are calculated for the spring period. Results are shown for forecasts started at (a) 0300, (b) 0900, (c) 1500, and (d) 2100 UTC. Only the cases where the RWS surface temperature was below 10°C were included in the forecast verification. The black line shows results for the control run without assimilated road surface temperatures, the red line for the control run with kriged and interpolated RWS temperatures, the blue line for the run using mobile observations, the brown line for the run using mobile observations with status-based correction, the pink line for the run using mobile observations with device-based correction, and the cyan line for the run using mobile observations with both status- and device-based correction.

forecast hours because the model had difficulties in the forecast of the daytime maximum values (Fig. 10b). The second control run has the smallest RMSE and the least negative bias values because the interpolated surface temperature values were warm enough to compensate for the otherwise too cold simulated temperatures. However, for the forecasts started at 1500 UTC the second control run gives the worst verification results (Fig. 10c). The reason for this was the too slowly cooling surface temperature in the model during evening, as the interpolated values made the already too warm model run even warmer.

Utilizing mobile observations clearly reduces the RMSE values at the beginning of the forecasts when

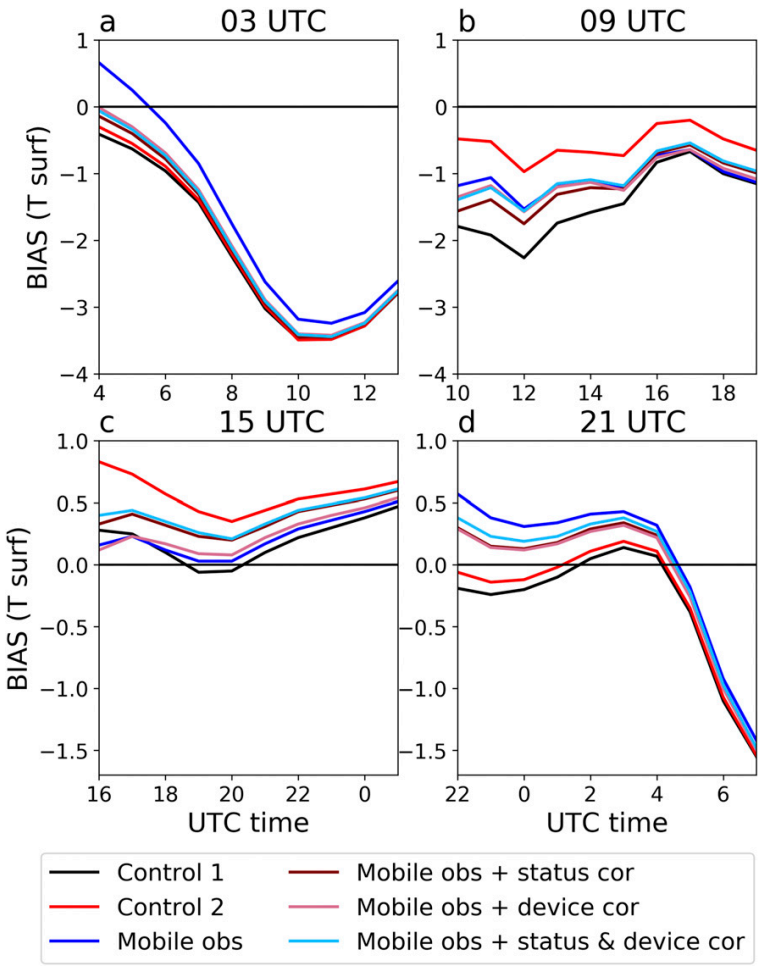

FIG. 11. As in Fig. 10, but for forecast bias.

compared to the first control run. However, the effects of the initial surface temperature corrections vary between forecast starting times. Using mobile observations caused a warm bias at the start of the model run for forecasts started at 0300 and 2100 UTC (Figs. 11a,d). The correction methods reduce this bias and also improve the RMSE values. However, in a few hours the model run without a statistical correction has the lowest RMSE values for the forecasts started at 0300 UTC (Fig. 10a). The initial warm bias compensates for the otherwise too cool modeled temperatures during morning and daytime. The second control run gives slightly greater RMSE values than the runs using mobile observations with statistical correction methods at the beginning of the forecasts started at 0300 UTC. In addition, the bias values of the second control run are slightly more negative. However, at the first forecast hour of the forecasts started at 2100 UTC the second control run gives clearly the smallest RMSE value (Fig. 10d). The bias value of the second control run is also nearest to zero of all the runs at the first forecast hour (Fig. 11d). Nonetheless, at the second forecast hour the RMSE value of the second control run is already at the same level as with the run using mobile observations with the status-based correction method.

The correction methods utilizing road status seem to somewhat increase the bias in the forecasts started at 1500 UTC (Fig. 11c), which was not seen in the verification results for the other forecast start times. These correction methods decrease the cold bias in warm temperatures according to Fig. 11. Indeed, the

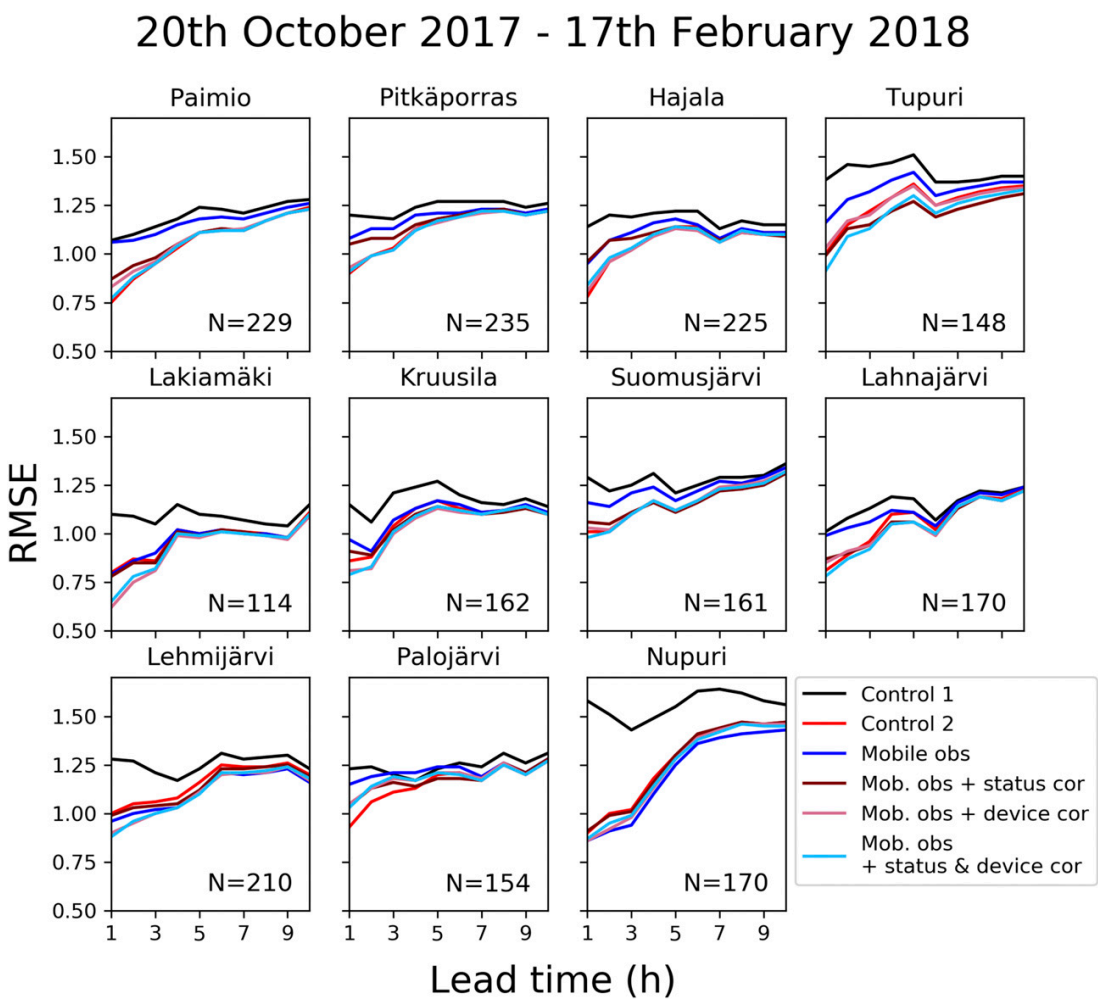

FIG. 12. RMSE values of road surface temperature forecasts calculated separately for each station. The forecasts were for the time period from 20 October 2017 to 17 February 2018. Only those stations that had 100 or more forecast cases with mobile observations available within three hours before the start of the forecast are included. The total amount of forecast cases for the station are shown in the bottom-right corner of each panel. Coordinates of the stations are shown in Table 3. Only the cases where the RWS surface temperature was below 10°C were included in the forecast verification The black line shows results for the control run without assimilated road surface temperatures, the red line for the control run with kriged and interpolated RWS temperatures, the blue line for the run using mobile observations, the brown line for the run using mobile observations with status-based correction, the pink line for the run using mobile observations with device-based correction, and the cyan line for the run using mobile observations with both status- and device-based correction.

correction methods bring the surface temperature closer to the observed values in the initialization phase of the 1500 UTC model run. However, the warmer temperatures in the initialization also cause the forecast part of the model run to be warmer. This increases the warm bias in the model as the simulated temperature cools down too slowly during the evening as mentioned above. However, the bias difference between the control run and the runs using road status in the correction method was not statistically significant with 95% confidence (Table A10). (The statistical significance between verification scores was calculated with the bootstrap method, and the results can be found in the appendix in Tables A4–A11.)

e. *Forecast verification for separate stations*

Verification results presented in Figs. 9–11 are calculated over all stations. The results for individual

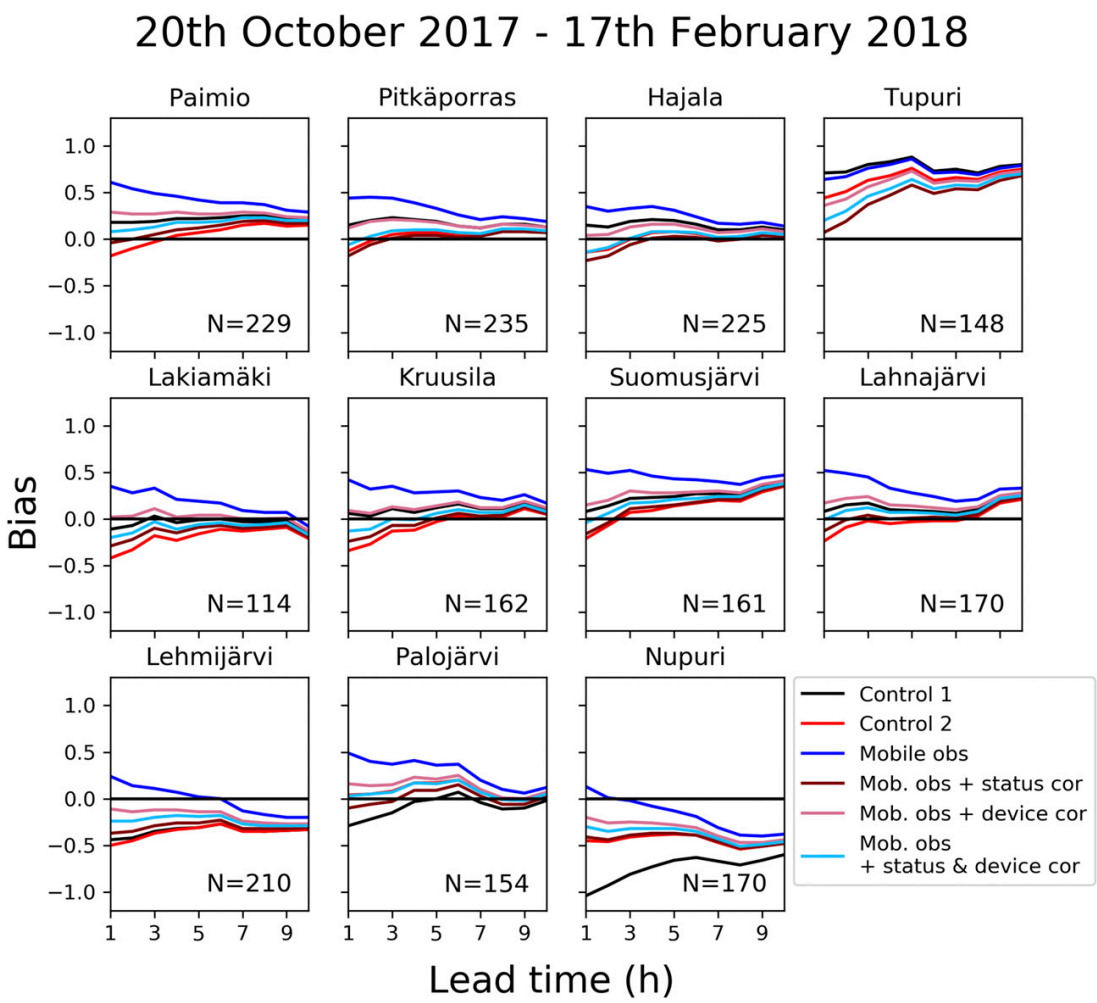

FIG. 13. As in Fig. 12, but for forecast bias.

stations vary due to geographical differences between station locations. As an example, Figs. 12 and 13 show RMSE and bias values calculated for separate stations during the winter period. Only those stations that had 100 or more forecast cases assimilating a mobile observation are included in the figures. The control run that did not assimilate any surface temperature observations gives the largest RMSE values at all stations, but the absolute values and the relative behavior compared to the other model runs vary between stations. The differences are biggest at Nupuri station, where the control run not using the surface temperature observations gives about 0.5°C higher RMSE values than the other runs at the first forecast hour. The control run without assimilated road surface temperatures had also strong negative bias at the start of the forecast for the Nupuri station. The Nupuri station is located on southeast–northwest-oriented motorway with two-lane roads in both directions. There is some forested area on both sides of the motorway, which has a warming effect during wintertime as the trees prevent longwave radiation emittance from the road to the atmosphere. The surface temperature at the station was often among the warmest of the stations used during the winter period. The FMI RWM assumes open-sky conditions, which is one reason for the cold bias in the forecasts. Assimilating mobile observations in the model reduced the cold bias and RMSE values for the Nupuri station.

TABLE A1. Explanations of model run names used in Tables A2–A11.

Control 1	Control run without mobile observations
Control 2	Control run with interpolated road surface temperature values
Mobile obs	Run using mobile observations
Status	Run using mobile observations with the status-based correction method
Device	Run using mobile observations with the device-based correction method
Stat+Dev	Run using mobile observations with the status- and device-based correction method

TABLE A2. Results of the statistical significance test for model RMSE differences during the midwinter period computed with the bootstrap method. Each row represents a different model run combination, and the columns represent forecast lead times. If the difference between the RMSE values of two model runs was found to be statistically significant with 95% confidence, the corresponding square is market with "x". The model run names are explained in Table A1.

	1	2	3	4	5	6	7	8	9	10
Control 1–Control 2	x	x	x	x	x	x	x	x	x	x
Control 1–Mobile obs	x	x	x	x	x	x	x	x	x	x
Control 1–Status	x	x	x	x	x	x	x	x	x	x
Control 1–Device	x	x	x	x	x	x	x	x	x	x
Control 1–Stat+Dev	x	x	x	x	x	x	x	x	x	x
Control 2–Mobile obs	x	x	x	x	x	x				
Control 2–Status	x				x				x	x
Control 2–Device					x	x				
Control 2–Stat+Dev			x	x	x	x	x		x	x
Mobile obs–Status	x	x	x	x	x	x	x	x	x	x
Mobile obs–Device	x	x	x	x	x	x	x	x	x	x
Mobile obs–Stat+Dev	x	x	x	x	x	x	x	x	x	x
Status–Device	x	x								
Status–Stat+Dev	x	x	x							
Device–Stat+Dev	x									

The model runs assimilating mobile observations without statistical correction gave the most positive bias values at most of the stations. However, at Tupuri station the bias of the control run that did not assimilate mobile observations is a little higher. The surface temperature measurements at the Tupuri station were often among the coldest when compared to the other stations included in the study during the winter period. The station is not located on the main motorway but on an older motorway leading from Helsinki to Turku. The smaller traffic amount might be one reason for the relative coldness of the station. Although the Teconer observations had a warm bias, the model run assimilating them without statistical correction still gave slightly smaller bias values than the control run without observations.

The relative RMSE and bias differences between the three model runs assimilating mobile observations with statistical correction vary between stations. At many stations the model run assimilating mobile observation with the status-based correction gives higher RSME values than the other two. It also has more negative bias at most of the stations. The relative accuracy of these model runs and model run assimilating the kriged surface temperature values also varies between stations.

The results for the spring period should be analyzed separately for each forecast start time as the daily temperature variation has a strong effect on the verification results. However, when verification results for individual stations were calculated for each forecast hour, the total amount of forecast cases at each station was rather low. Not including the cases where the RWS surface temperature was above 10°C also affected the total number of cases. Because of the small sample size, the results are not considered reliable enough to be analyzed further in this paper.

5. Discussion and conclusions

Vehicle-based observations provide a cost-effective way for road condition monitoring on a dense spatial scale. Observations of initial road conditions are important, because even small variations can change the road condition forecast especially when the road surface temperature is near the freezing point. However, the quality of mobile observations and their effect on the forecast accuracy must be studied before operational implementation.

The effect of mobile observations on road weather forecasts was studied by including mobile surface temperature observations in the road weather model initialization. Mobile observations improved the accuracy of the road surface temperature forecasts when compared to the forecast scenario in which there would not be road weather stations in the area. Adding statistical correction to the measurements increased the accuracy further during the winter period. However, there was no clear improvement for the winter period when the results were compared to the forecast scenario with a dense RWS network. The relative differences between runs had more variation during the spring period when the daily surface temperature variation was stronger. The best performing statistical correction method of the three tested methods seemed to be the one that was dependent on the road status and the individual Teconer device when the corrected values were compared directly to the RWS observations. However, a clear winner among the correction methods could not be found when the forecast verification results were compared with each other, as the effects of the correction on the verification results are complex. For the winter period, the

TABLE A3. As in Table A2, but for bias values.

	1	2	3	4	5	6	7	8	9	10
Control 1–Control 2	x	x	x	x	x	x	x	x	x	x
Control 1–Mobile obs	x	x	x	x	x	x	x	x	x	x
Control 1–Status	x	x	x	x	x	x	x	x	x	x
Control 1–Device	x	x	x	x	x	x	x	x	x	x
Control 1–Stat+Dev										
Control 2–Mobile obs	x	x	x	x	x	x	x	x	x	x
Control 2–Status										
Control 2–Device	x	x	x	x	x	x	x	x	x	x
Control 2–Stat+Dev	x	x	x	x	x	x	x	x	x	x
Mobile obs–Status	x	x	x	x	x	x	x	x	x	x
Mobile obs–Device	x	x	x	x	x	x	x	x	x	x
Mobile obs–Stat+Dev	x	x	x	x	x	x	x	x	x	x
Status–Device	x	x	x	x	x	x	x	x	x	x
Status–Stat+Dev	x	x	x	x	x	x	x	x	x	x
Device–Stat+Dev	x	x	x	x	x	x	x	x	x	x

TABLE A5. As in Table A2, but for RMSE values of forecasts started at 0900 UTC during the spring period.

	1	2	3	4	5	6	7	8	9	10
Control 1–Control 2	x	x	x	x	x	x	x		x	x
Control 1–Mobile obs	x		x	x				x		
Control 1–Status			x						x	x
Control 1–Device			x					x		x
Control 1–Stat+Dev			x						x	x
Control 2–Mobile obs		x	x			x	x	x	x	x
Control 2–Status	x	x	x	x	x	x	x		x	x
Control 2–Device	x	x	x		x	x	x	x	x	x
Control 2–Stat+Dev	x	x	x	x	x	x	x		x	x
Mobile obs–Status	x	x	x	x					x	x
Mobile obs–Device	x	x			x	x			x	x
Mobile obs–Stat+Dev	x	x							x	x
Status–Device				x			x		x	x
Status–Stat+Dev			x							
Device–Stat+Dev							x		x	x

error in statistically corrected mobile observations and kriged values was about the same order of magnitude as the forecast error. However, the surface temperature observation error in these cases was not found to be the dominant factor in the forecast error but the other modeling errors seemed to have a strong effect on the forecast accuracy.

The verification results were affected by the asphalt sensor overheating, which may have caused too pessimistic results during the spring period. In addition, using the average of mobile measurements done within 50 m of the station does not always correspond well to the RWS measurements done at a certain spot. The mobile measurements might have been done for example in a different lane, and the road surface temperature can vary within this distance especially if the road is partially in shadow. This might have a negative effect on the accuracy of the model runs assimilating mobile observations, as the RWS observations are used in the verification. The results of this study still provide useful information about the effects of the mobile observations and correction methods used on the road surface temperature forecasts.

An interesting topic for further studies would be to test how many stations are needed to get the same forecast accuracy as with mobile observations. This could be done by performing the road surface temperature kriging analysis with different amounts of observation points and using the values in RWM initialization. The studied area had only small elevation variations, and the effect of the mobile observations would probably have been greater in the areas with a more undulating landscape. In addition, the verification was performed only on the road weather station points. Along the roads there can be points, such as bridges, which are more prone to freezing and would be more easily monitored with mobile measurements.

TABLE A4. As in Table A2, but for RMSE values of forecasts started at 0300 UTC during the spring period.

	1	2	3	4	5	6	7	8	9	10
Control 1–Control 2	x	x	x	x	x					
Control 1–Mobile obs	x	x	x	x	x	x	x	x	x	x
Control 1–Status	x	x	x	x	x	x	x	x		x
Control 1–Device	x	x	x	x	x	x		x		x
Control 1–Stat+Dev	x	x	x	x	x	x	x			x
Control 2–Mobile obs	x		x	x	x	x	x	x	x	x
Control 2–Status		x	x	x	x	x	x	x		
Control 2–Device		x	x	x	x	x	x			
Control 2–Stat+Dev		x	x	x	x	x	x		x	
Mobile obs–Status	x		x	x	x	x	x	x	x	x
Mobile obs–Device	x	x	x	x	x	x	x	x	x	x
Mobile obs–Stat+Dev	x		x	x	x	x	x	x	x	x
Status–Device										
Status–Stat+Dev										
Device–Stat+Dev										

TABLE A6. As in Table A2, but for RMSE values of forecasts started at 1500 UTC during the spring period.

	1	2	3	4	5	6	7	8	9	10
Control 1–Control 2	x	x	x					x	x	x
Control 1–Mobile obs		x								
Control 1–Status										
Control 1–Device		x	x	x	x					
Control 1–Stat+Dev										
Control 2–Mobile obs	x	x	x	x	x	x	x	x	x	x
Control 2–Status				x	x	x	x	x	x	x
Control 2–Device	x	x	x	x	x	x	x	x	x	x
Control 2–Stat+Dev		x	x	x	x	x	x	x	x	x
Mobile obs–Status	x	x	x							
Mobile obs–Device			x	x	x	x	x			
Mobile obs–Stat+Dev	x	x								
Status–Device	x	x	x	x	x		x	x	x	x
Status–Stat+Dev		x	x		x	x				
Device–Stat+Dev	x	x	x	x						

TABLE A7. As in Table A2, but for RMSE values of forecasts started at 2100 UTC during the spring period.

	1	2	3	4	5	6	7	8	9	10
Control 1–Control 2	x	x	x	x	x				x	x
Control 1–Mobile obs		x	x					x	x	x
Control 1–Status	x	x	x	x				x	x	x
Control 1–Device		x	x	x				x	x	x
Control 1–Stat+Dev		x	x	x				x	x	x
Control 2–Mobile obs	x	x				x	x		x	x
Control 2–Status	x								x	x
Control 2–Device	x					x			x	x
Control 2–Stat+Dev	x				x	x	x		x	x
Mobile obs–Status	x	x	x	x						x
Mobile obs–Device	x								x	x
Mobile obs–Stat+Dev										x
Status–Device			x	x						
Status–Stat+Dev	x	x	x	x	x	x	x			
Device–Stat+Dev									x	

TABLE A9. As in Table A2, but for bias values of forecasts started at 0900 UTC during the spring period.

	1	2	3	4	5	6	7	8	9	10
Control 1–Control 2	x	x	x	x	x	x	x	x	x	x
Control 1–Car obs	x	x	x	x	x	x	x			
Control 1–Status		x	x	x	x	x	x	x	x	x
Control 1–Device	x	x	x	x	x	x			x	x
Control 1–Stat+Dev	x	x	x	x	x	x	x	x	x	x
Control 2–Car obs	x	x	x	x	x	x	x	x	x	x
Control 2–Status	x	x	x	x	x	x	x	x	x	x
Control 2–Device	x	x	x	x	x	x	x	x	x	x
Control 2–tat+Dev	x	x	x	x	x	x	x	x	x	x
Car obs–Status	x	x	x	x	x			x	x	x
Car obs–Device	x					x			x	x
Car obs–Stat+Dev	x						x	x	x	x
Status–Device	x	x	x	x	x		x	x	x	x
Status–Stat+Dev	x	x	x	x	x	x	x	x	x	x
Device–Stat+Dev						x	x	x	x	x

An interesting topic for future research would also be to repeat the study in an area with only a few road weather stations such as the northern part of Finland. This study could utilize measurements from a truck fleet that has been equipped with measurement devices within the EU-funded Intelligent Arctic Trucks project (Sukuvaara et al. 2018).

The mobile surface temperature observations included in this study were done with just one sensor type. However, as the amount of available mobile observations increases, further studies should be made to assess their effect on the road weather forecasts. Different kinds of statistical correction methods will be needed as other sensors might behave differently, and road status measurements are not always available. In addition, many vehicles do not measure road surface temperature. It needs to be studied further if more commonly measured variables, like air temperature, can be used to improve road weather forecast accuracy. Air temperature measurements cannot be used in the model in the same way as road surface temperature observations were used in this study to adjust the radiation fluxes. However, they might be more usable when implemented in 3D numerical weather prediction models. There is already an ongoing project that aims to implement vehicle observations in a regional operational weather prediction model in Germany (Riede et al. 2018).

Acknowledgments. We express our gratitude to Prof. Heikki Järvinen (University of Helsinki), Dr. Timo Sukuvaara (FMI), and M.Sc. Petteri Karsisto (FMI) for their valuable comments and suggestions. The Finnish

TABLE A8. As in Table A2, but for bias values of forecasts started at 0300 UTC during the spring period.

	1	2	3	4	5	6	7	8	9	10
Control 1–Control 2										
Control 1–Car obs	x	x	x	x	x	x	x	x	x	x
Control 1–Status	x	x	x	x	x	x				
Control 1–Device	x	x	x	x	x	x	x	x	x	x
Control 1–Stat+Dev	x	x	x	x	x	x	x	x	x	x
Control 2–Car obs	x	x	x	x	x	x	x	x	x	x
Control 2–Status	x	x	x	x	x	x	x	x	x	
Control 2–Device	x	x	x	x	x	x	x	x	x	x
Control 2–Stat+Dev	x	x	x	x	x	x	x	x	x	x
Car obs–Status	x	x	x	x	x	x	x	x	x	x
Car obs–Device	x	x	x	x	x	x	x	x	x	x
Car obs–Stat+Dev	x	x	x	x	x	x	x	x	x	x
Status–Device	x	x	x	x	x	x	x	x		
Status–Stat+Dev	x	x	x	x	x	x	x			
Device–Stat+Dev	x	x	x	x	x					

TABLE A10. As in Table A2, but for bias values of forecasts started at 1500 UTC during the spring period.

	1	2	3	4	5	6	7	8	9	10
Control 1–Control 2	x	x	x	x	x	x	x	x	x	x
Control 1–Car obs				x	x	x	x	x	x	x
Control 1–Status		x	x	x	x	x	x	x	x	x
Control 1–Device				x	x	x	x	x	x	x
Control 1–Stat+Dev		x	x	x	x	x	x	x	x	x
Control 2–Car obs	x	x	x	x	x	x	x	x	x	x
Control 2–Status	x	x	x	x	x	x	x	x	x	x
Control 2–Device	x	x	x	x	x	x	x	x	x	x
Control 2–Stat+Dev	x	x	x	x	x	x	x	x	x	x
Car obs–Status	x	x	x	x	x	x	x	x	x	x
Car obs–Device			x	x	x	x	x	x	x	x
Car obs–Stat+Dev	x	x	x	x	x	x	x	x	x	x
Status–Device	x	x	x	x	x	x	x	x	x	x
Status–Stat+Dev	x									
Device–Stat+Dev	x	x	x	x	x	x	x	x	x	x

TABLE A11. As in Table A2, but for bias values of forecasts started at 2100 UTC during the spring period.

	1	2	3	4	5	6	7	8	9	10
Control 1–Control 2	x	x	x	x	x		x	x	x	
Control 1–Car obs	x	x	x	x	x	x	x	x	x	x
Control 1–Status	x	x	x	x	x	x	x	x	x	x
Control 1–Device	x	x	x	x	x	x	x	x	x	x
Control 1–Stat+Dev	x	x	x	x	x	x	x	x	x	x
Control 2–Car obs	x	x	x	x	x	x	x	x	x	x
Control 2–Status	x	x	x	x	x	x	x	x	x	x
Control 2–Device	x	x	x	x	x	x	x	x	x	x
Control 2–Stat+Dev	x	x	x	x	x	x	x	x	x	x
Car obs–Status	x	x	x	x	x	x	x	x	x	x
Car obs–Device	x	x	x	x	x	x	x	x	x	x
Car obs–Stat+Dev	x	x	x	x	x	x	x	x	x	x
Status–Device									x	
Status–Stat+Dev	x	x	x	x	x	x	x	x		
Device–Stat+Dev	x	x	x	x	x	x	x	x	x	

Meteorological institute and the University of Oulu have participated in several research projects that supported this work and the authors gratefully acknowledge the support provided. The projects were Intelligent Arctic trucks and Sod5G funded by European Regional Development Fund of the European Union and Regional Council of Lapland, WiRMa funded by Interreg Nord fund of European Union and Regional Council of Lapland, 5G-Safe funded by Business Finland, 6Genesis Flagship funded by Academy of Finland, and the AI Enhanced Mobile Edge Computing project funded by the Future Makers program of Jane and Aatos Erkko Foundation and Technology Industries of Finland Centennial Foundation. We want to thank also all financiers and partners who participated in these projects.

APPENDIX

Statistical Significance of the Model Bias and RMSE Differences

The results of statistical significance tests performed with the bootstrap method are presented in this appendix. Table A1 provides explanations for the model run names used in Tables A2–A11. Tables A2 and A3 show the results of RSME and bias differences for each model run pair during the winter period. Tables A4–A11 show the corresponding results for the spring period, except the results are separated by the forecast start time.

REFERENCES

Aalto, J., P. Pirinen, J. Heikkinen, and A. Venäläinen, 2013: Spatial interpolation of monthly climate data for Finland: Comparing the performance of kriging and generalized additive models. *Theor. Appl. Climatol.*, **112**, 99–111, https://doi.org/10.1007/s00704-012-0716-9.

Allegretti, M., and S. Bertoldo, 2014: Cars as a diffuse network of road-environment monitoring nodes. *Wireless Sensor Network*, **6**, 184–191, https://doi.org/10.4236/wsn.2014.69018.

Bengtsson, L., and Coauthors, 2017: The HARMONIE–AROME model configuration in the ALADIN–HIRLAM NWP system. *Mon. Wea. Rev.*, **145**, 1919–1935, https://doi.org/10.1175/MWR-D-16-0417.1.

Bogren, J., T. Gustavsson, M. Karlsson, and U. Postgård, 2000: The impact of screening on road surface temperature. *Meteor. Appl.*, **7**, 97–104, https://doi.org/10.1017/S135048270000150X.

Brutsaert, W., 1984: *Evaporation into the Atmosphere: Theory, History, and Applications*. D. Reidel Publishing Company, 299 pp.

Chapman, L., and J. E. Thornes, 2006: A geomatics-based road surface temperature prediction model. *Sci. Total Environ.*, **360**, 68–80, https://doi.org/10.1016/j.scitotenv.2005.08.025.

Cressie, N., 1993: *Statistics for Spatial Data*. John Wiley & Sons, 900 pp.

Crevier, L. P., and Y. Delage, 2001: A new model for road-condition forecasting in Canada. *J. Appl. Meteor.*, **40**, 2026–2037, https://doi.org/10.1175/1520-0450(2001)040<2026:MANMFR>2.0.CO;2.

Efron, B., and R. J. Tibshirani, 1993: *An Introduction to the Bootstrap*. Chapman and Hall, 436 pp.

Gustavsson, T., 1999: Thermal mapping—A technique for road climatological studies. *Meteor. Appl.*, **6**, 385–394, https://doi.org/10.1017/S1350482799001334.

Haavasoja, T., J. Nylander, and P. Nylander, 2012: Experiences of Mobile Road Condition Monitoring. *16th Int. Road Weather Conf. (SIRWEC)*, Helsinki, Finland, SIRWEC, http://www.teconer.fi/downloads/SIRWEC_2012_Experiences%20of%20Mobile%20RCM_ID14.pdf.

Hogan, R. J., and I. B. Mason, 2011: Deterministic forecasts of binary events. *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, 2nd ed. I. T. Jolliffe and D. B. Stephenson, Eds., John Wiley and Sons, 31–59, https://doi.org/10.1002/9781119960003.ch3.

Kangas, M., M. Heikinheimo, and M. Hippi, 2015: RoadSurf—A modelling system for predicting road weather and road surface conditions. *Meteor. Appl.*, **22**, 544–533, https://doi.org/10.1002/met.1486.

Karsisto V., 2018: Road surface temperature forecast study HKI-TKU 0708. Zenodo.org, accessed 25 September 2018, https://doi.org/10.5281/zenodo.1434636.

——, P. Nurmi, M. Kangas, M. Hippi, C. Fortelius, S. Niemelä, and H. Järvinen, 2016: Improving road weather model forecasts by adjusting the radiation input. *Meteor. Appl.*, **23**, 503–513, https://doi.org/10.1002/met.1574.

——, S. Tijm, and P. Nurmi, 2017: Comparing the performance of two road weather models in the Netherlands. *Wea. Forecasting*, **32**, 991–1006, https://doi.org/10.1175/WAF-D-16-0158.1.

Lovén, L., V. Karsisto, H. Järvinen, M. J. Sillanpää, T. Leppänen, E. Peltonen, S. Pirttikangas, and J. Riekki, 2019: Mobile road weather sensor calibration by sensor fusion and linear mixed models. *PLoS ONE*, **14**, e0211702 https://doi.org/10.1371/journal.pone.0211702.

Patankar, S. V., 1980: *Numerical Heat Transfer and Fluid Flow*. McGraw-Hill, 197 pp.

Petty, K. R., and W. P. Mahoney, 2007: Weather applications and products enabled through vehicle infrastructure integration (VII). U.S Department of Transportation, 124 pp.

Riede, H., A. Bouras, Z. Paschalidi, J. W. Acevedo-Valenciaa, T. Kratzsch, and J. Nachtigall, 2018: Flowkar: Using high-resolution data from vehicle sensors to improve operational weather products. *19th Conf. of Standing International Road Weather Commission (SIRWEC)*, Smolenice, Slovakia, SIRWEC, 95–96.

Robinson, G. K., 1991: That BLUP is a good think: The estimation of random effects. *Stat. Sci.*, **6**, 15–51, https://doi.org/10.1214/ss/1177011926.

Shao, J., P. J. Lister, G. D. Hart, and H. B. Pearson, 1996: Thermal mapping: Reliability and repeatability. *Meteor. Appl.*, **3**, 325–330, https://doi.org/10.1002/met.5060030405.

——, J. C. Swanson, R. Patterson, P. J. Lister, and A. N. McDonald, 1997: Variation of winter road surface temperature due to topography and application of thermal mapping. *Meteor. Appl.*, **4**, 131–137, https://doi.org/10.1017/S135048279700042X.

Sukuvaara, T., K. Mäenpää, S. Kantomaa, D. Stepanova, M. Hippi, and V. Karsisto, 2018: Intelligent traffic enabled advanced road weather infrastructures in Arctic conditions. *European Geosciences Union General Assembly 2018*, Vol. 20, EGU2018-95, Vienna, Austria, European Geosciences Union, https://meetingorganizer.copernicus.org/EGU2018/EGU2018-95.pdf.

Thornes, J. E., 1991: Thermal mapping and road-weather information systems for highway engineers. *Highway Meteorology*, A. H. Perry and L. J. Symons, Eds., E & FN Spon, 39–67.

Vaisala, 2001: Vaisala Road/runway surface and depth sensor DRS511. Tech. Note, 2 pp., https://www.vaisala.com/sites/default/files/documents/WEA-RDS-G-DRS511%20Datasheet.pdf.

West, B. T., K. B. Welch, and A. T. Galecki, 2007: *Linear Mixed Models: A Practical Guide Using Statistical Software*. Chapman & Hall/CRC, 374 pp., https://doi.org/10.1002/sim.3167.

ILMATIETEEN LAITOS
METEOROLOGISKA INSTITUTET
FINNISH METEOROLOGICAL INSTITUTE

FINNISH METEOROLOGICAL INSTITUTE
Erik Palménin aukio 1
P.O. Box 503
FI-00560 HELSINKI
tel. +358 29 539 1000
WWW.FMI.FI

FINNISH METEOROLOGICAL INSTITUTE
CONTRIBUTIONS No. 153
ISSN 0782-6177
ISBN 978-952-336-078-5 (paperback)
ISBN 978-952-336-079-2 (pdf)
https://doi.org/10.35614/isbn.9789523360792

Edita Prima Oy
Helsinki 2019